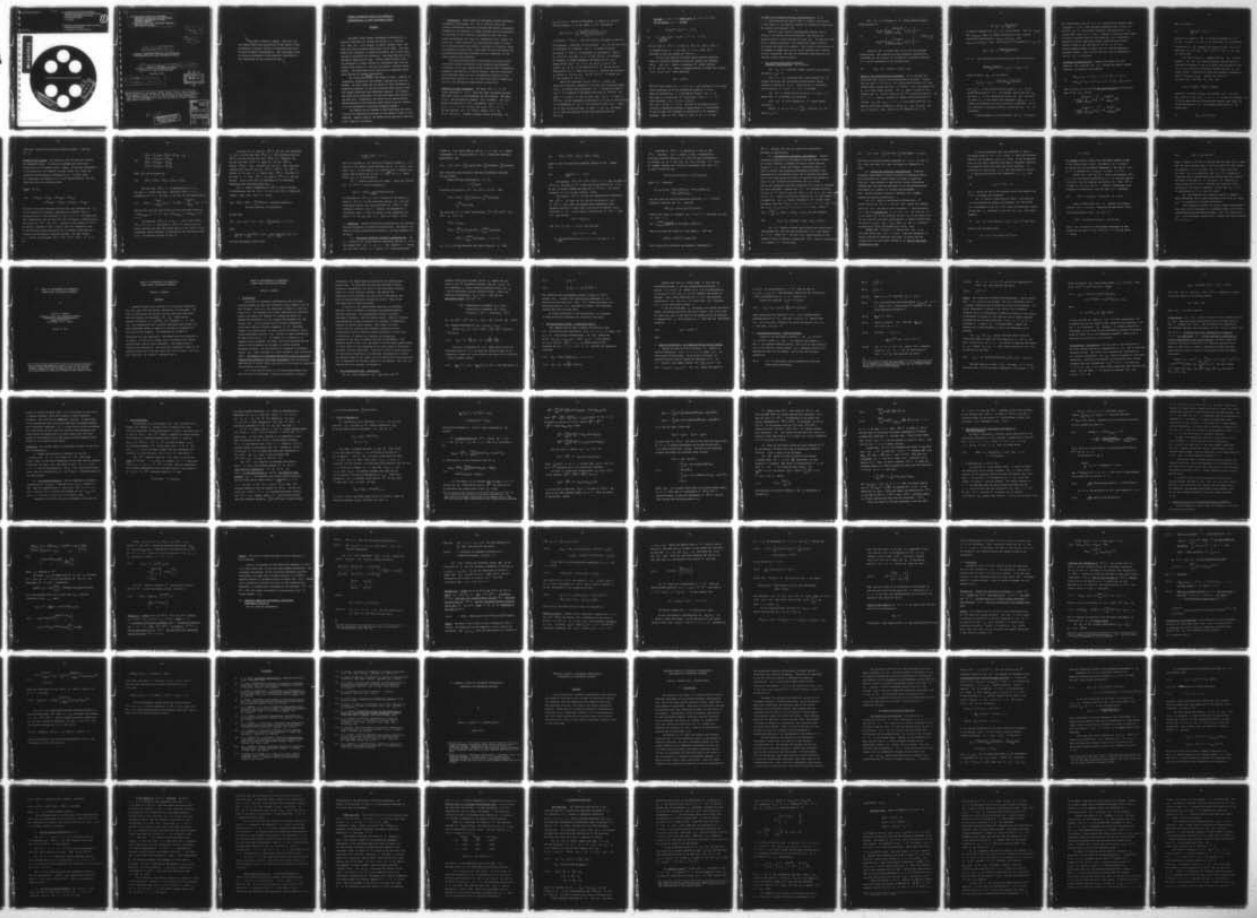
AD-A032 824 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
I, II CONVERGENCE AND RATE OF CONVERGENCE THEOREMS FOR CONSTRAI--ETC(U)
AUG 76 H J KUSHNER, S LAKSHMIVARAHAN N00014-76-C-0279

UNCLASSIFIED LCDS-TR-76-1 NL



LEFSCHETZ CENTER FOR DYNAMICAL SYSTEMS

(6) I,II Convergence and Rate of Convergence Theorems for Constrained and Unconstrained Stochastic Approximation, via Weak Convergence methods.
III Numerical Studies for Constrained Stochastic Approximation Problems,

(11) Aug 76

(12) 139 p.

(14) LCDS-TR-76-I

I. GENERAL CONVERGENCE RESULTS FOR STOCHASTIC APPROXIMATIONS VIA WEAK CONVERGENCE THEORY +

(10) Brown Univ.
Harold J. Kushner S. Lakshmivarahan
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics

December 1975

(15) N0014-76-C-0279,
AF-AFOSR-2078C-71

+Brown University, Providence, Rhode Island 02912. This research was supported in part by the Air Force Office of Scientific Research under AFOSR-71-2078C, in part by the National Science Foundation under Eng-73-03846-~~A01~~ and in part by the Office of Naval Research under N000-14-76-C-0279.

This report contains 3 papers. The first two (by Kushner) deal with applications of the theory of weak convergence to general results in constrained and unconstrained stochastic approximation. The third paper gives an extensive discussion of the numerical properties of four algorithms for the constrained case.

GENERAL CONVERGENCE RESULTS FOR STOCHASTIC APPROXIMATIONS via WEAK CONVERGENCE THEORY

ABSTRACT

The paper treats general convergence conditions for a class of algorithms for finding the minima of a function $f(x)$ when $f(x)$ is of unknown (or partly unknown) form - and when only noise corrupted observations can be taken. Such problems occur frequently in adaptive processes, and in many applications to statistics and estimation. The algorithms are of the stochastic approximation type. Several forms are dealt with - for estimation in either discrete or continuous time, with and without side constraints, and with or without - periodic search renewal. The algorithms can be considered as sequential Monte Carlo methods for systems optimization.

The innovations partly concern the method of proof. However, an interesting 'constrained' and'renewed' algorithm are also considered. By using ideas from the theory of weak convergence of probability measures, we can get relatively short proofs, under much weaker conditions than heretofore required. For example, the noise can be correlated, and there are fewer restrictions on the step size. Furthermore, the nature of the method permits generalizations to more abstract cases (which occur for example, if we are optimizing a distributed parameter system).

The results can be extended in many directions and variations of the technique can be used to get bounds on rates of convergence. Special forms of the method can be applied to many well known 'adaptive' procedures.

1. Introduction. Using results in the theory of weak convergence of measures (Billingsley [1]), and in stability theory for ordinary differential equations, we prove some general convergence theorems for the sequences of random variables which are generated by algorithms of the stochastic approximation type (see e.g., Kiefer and Wolfowitz [6], Blum [2], Schmetterer [15]), Fabian [5], Wasan [16]. Such algorithms are used when one wishes to locate, via a recursive Monte-Carlo method, a minimum of a function, under the handicap of "noisy" data. Algorithms for both constrained and unconstrained optimization problems will be considered, and for rather general noise processes.

The noise processes that we allow seem to be more typical (than those usually considered in the stochastic approximation literature) of the type usually found in applications to sequential parametric optimization of control systems. Some work concerning noise processes where the usual orthogonality condition is not satisfied was reported in [3], but under very special assumptions, and for a one dimensional Robbins-Munro process only.

Remarks on a weak convergence. The space $D^m(-\infty,\infty)$ of R^m valued functions on $(-\infty,\infty)$ which are right continuous and have left hand limits, is most convenient for our purposes. The space $D^1[0,\infty)$ was discussed by Lindvall [11] in detail. Let $g_j(\cdot)$ denote the function on $[0,\infty)$ defined by: $g_j(t) = 1$, $t \leq j$, $g_j(t) = j + 1 - t$ on $[j,j+1]$, and $g_j(t) = 0$, $t \geq j + 1$. For any function $x(\cdot)$ in $D^1[0,\infty)$, define $x^j(\cdot)$ by $x^j(t) = x(t)g_j(t)$. Lindvall proceeds roughly as follows. Let

$d_{0,j}(x(\cdot),y(\cdot))$ denote the Skorokhod d_0 metric on $D[0,j+1]$ (See Billingsley p. 111 on, where $j = 1$). The metric

$$d(x(\cdot),y(\cdot)) = \sum_{j=1}^{\infty} \frac{d_{0,j}(x^j(\cdot),y^j(\cdot))}{1 + d_{0,j}(x(\cdot)^j,y(\cdot)^j)} 2^{-j}$$

is defined on $D^1[0,\infty)$. Under this metric, the space is complete and separable. Tightness (see Billingsley [1] for the definition of tightness) of a sequence $\{X_n(\cdot)\}$ of random functions with paths in $D^1[0,\infty)$ is equivalent to tightness of $\{X_n^j(\cdot)\}$ on $D^1[0,j+1]$, for each $j \geq 1$. To get $D^1(-\infty,\infty)$, as indicated by Lindvall, we simply symmetrize the definition of $g_j(\cdot)$ Simply define $\bar{g}_j(\cdot)$ on $(-\infty,\infty)$ by $\bar{g}_j(t) = g_j(|t|)$ and replace g_j by $\bar{g}_j$ in the definition of x^j, and we use the above metric $d(\cdot,\cdot)$, where now $d_{0,j}(\cdot,\cdot)$ denotes the Skorokhod d_0 metric on $D^1[-j-1,j+1]$. We use $D^m(-\infty,\infty)$ to denote the m fold product of $D^1(-\infty,\infty)$.

While our processes will be $D^m(-\infty,\infty)$ valued, the limits will generally (except for one case) have continuous paths w.p.1. It follows from Billingsley [1] , Theorem 15.5, that, if $\{X_n(\cdot)\}$ is a sequence of processes with paths in $D^m[-T,T]$ for some $T > 0$, and if (1) and (2) hold, then $\{X_n(\cdot)\}$ is tight on $D^m[-T,T]$ and any separable process which is a limit in distribution of a subsequence of $\{X_n(\cdot)\}$ has continuous paths w.p.1. Hence, if (1) and (2) hold for each T, where δ and n_0 can depend on T, then the assertion of the last sentence holds for $D^m(-\infty,\infty)$.

<u>For each</u> $\eta > 0$, $\varepsilon > 0$, <u>there is an</u> $N_\eta < \infty$, a $\delta \in (0,1)$ <u>and an integer</u> $n_0 < \infty$ <u>so that</u>

$$P\{|X_n(0)| \geq N_\eta\} \leq \eta, \quad n \geq 1 \tag{1}$$

$$P\{\sup_{\substack{|t-s|\leq\delta \\ T\geq t\geq s\geq -T}} |X_n(t) - X_n(s)| \geq \varepsilon\} \leq \eta, \quad n \geq n_0. \tag{2}$$

We note that if $\{X^n(\cdot)\}$ is tight on $D^m(-\infty,\infty)$, and if $X^n(t) \to 0$ in probability as $n \to \infty$, for each $t \in (-\infty,\infty)$, then $X^n(\cdot)$ converges to the zero element of $D^m(-\infty,\infty)$.

We will also need the following result from the stability theory of ordinary differential equations (LaSalle and Lefschetz [10]). Let $g(\cdot)$ denote a continuous R^r valued function on R^r and let $X(\cdot)$ denote a bounded function on either the interval $(-\infty,\infty)$ or on $[0,\infty)$ which satisfies

$$\dot{X}(t) = g(X(t)).$$

Then all limit points (as $t \to \infty$) of $X(\cdot)$ are contained in the largest finite invariant set of the differential equation $\dot{y} = g(y)$, where an invariant set M is defined as follows. Let $y \in M$, then there is a continuous R^r valued path $\{y(t), \infty > t > -\infty\}$, which satisfies $y(0) = y$, $\dot{y}(t) = g(y(t))$, and $y(t) \in M$, all $t \in (-\infty,\infty)$. The use of the doubly infinite interval is especially valuable in applications - it facilitates, as will be seen, the treatment of stationarity or of asymptotic problems. Also, if $X(t)$ tends to a set N as $t \to \infty$, then

it tends to the largest invariant set contained in N, as $t \to \infty$ (which may be much smaller than N itself). This result is very useful in the stability analysis of differential equations. See LaSalle and Lefschetz [10].

Section 2 will treat an unconstrained problem, and in Section 3, we discuss several extensions. The results are generalizations of those obtained by Ljung [12]. The approach taken here is somewhat different. Our method uses rather different techniques and seems to be simpler. Also, owing to the features of the method, it seems that results can be obtained for a greater variety of problems, and also for stochastic approximations with abstract valued random variables.

2. The Unconstrained Problem Classical Stochastic Approximation. Let us assume

(A1) $\{a_n\}$ is a (possibly random) positive null sequence satisfying $\sum_n a_n = \infty$.

(A2) $f(\cdot)$ is a continuous real valued function on R^r, bounded from below and continuously differentiable (whose gradient is denoted by $f_x(\cdot)$). Each solution to the differential equation $\dot{x} = -f_x(x)$ on $[0,\infty)$ is bounded.

(A3) Let S denote the set of points $\{x: f_x(x) = 0\}$. Assume that S is bounded, and connected.

(A4) $\{\beta_n\}$ is a null sequence of R^r valued random variables.

Define t_n by $t_0 = 0$, $t_n = \sum_{i=0}^{n-1} a_i$ and write m_t for $\max\{n: t_n \leq t\}$.

(A5) $\{\xi_n\}$ is a sequence of R^r valued random variables which satisfy (3).

$$\lim_N \sup P\left\{\max_{0\le t\le T}\left|\sum_{i=m(t_N)}^{i=m(t_N+t)-1} a_i\xi_i\right| \ge \varepsilon\right\} = 0$$

$$\lim_N \sup P\left\{\max_{-T\le t\le 0}\left|\sum_{i=m(t_N+t)}^{m(t_N)-1} a_i\xi_i\right| \ge \varepsilon\right\} = 0 \qquad (3)$$

, each $T > 0$, $\varepsilon > 0$.

Condition (A5) is rather weak, and it will be discussed further below. The iteration for many stochastic approximation algorithms for the location of a minimum of a regression function $f(\cdot)$ can be represented in the following form

$$X_{n+1} = X_n - a_n f_x(X_n) + a_n\beta_n + a_n\xi_n. \qquad (4)$$

Remark on the Kiefer-Wolfowitz procedure. Let us consider one particular case, that of the Kiefer-Wolfowitz procedure. Suppose that for each $x \in R^r$, $H(\cdot|x)$ is a distribution function of a real valued random variable and that $\int y\, H(dy|x) = f(x)$, and suppose that X_0 is a given random variable. Let c_n denote a positive null sequence and e_i the unit vector in the i^{th} coordinate direction. Let Y_j, $j = 1,\ldots,$ denote a sequence of random variables obtained as follows. Y_{2i} and Y_{2i-1}, $i = 1,\ldots,r$, are random draws with law (regular conditional probability, given X_0). $H_{2i}(\cdot|X_0+e_ic_0)$ and $H_{2i-1}(\cdot|X_0-e_ic_0)$, resp. Define the i^{th} component of X_1 by

$$X_1^i = X_0^i - a_n \frac{[Y_{2i} - Y_{2i-1}]}{2c_0} .$$

In general, suppose that $X_0, \ldots, X_n$ are available. Then let Y_{2n+2i} and $Y_{2n+2i-1}$, resp., be random variables with laws (regular conditional distributions) $H_{2n+2i}(\cdot | X_n + e_i c_n)$ and $H_{2n+2i-1}(\cdot | X_n - e_i c_n)$, resp., and define the i^{th} component of X_{n+1} by

$$X_{n+1}^i = X_n^i - a_n \frac{[Y_{2n+2i} - Y_{2n+2i-1}]}{2c_n} .$$

If $f(\cdot)$ has continuous second derivatives, then we can write

$$\frac{Y_{2n+2i} - Y_{2n+2i-1}}{2c_n} = f_{x_i}(X_n) + c_n \beta'_{n,i} + \xi_n ,$$

where we define $\beta'_{n,i}$ by the equality

$$f_{x_i}(X_n) + c_n \beta_{n,i} = \frac{f(X_n + e_i c_n) - f(X_n - e_i c_n)}{2c_n} ,$$

and $f_{x_i}(\cdot)$ *is the derivative with respect to* x_i. *If the* second derivatives of $f(\cdot)$ are bounded, then so are the $\beta'_{n,i}$ The term ξ_n is the "observation" noise, and $\{X_n\}$ satisfies (4) where β_n is the vector with components $c_n \beta'_{n,1}, \ldots$. Of course, the recursion (4) can be obtained in many ways. We do not, contrary to the usual practice, assume that the $\{\xi_i\}$ are orthogonal.

In many examples, the distribution $H_n(\cdot | x)$ is replaced

by a distribution $H_n'(\cdot|x)$ of a R^r valued random variable, where $\int yH_n'(dy|x) = f_x(x)$. Then the Kiefer-Wolfowitz form is still (4), but with $\beta_n = 0$. With either regression, the definition of ξ_n implies that $E[\xi_n | X_n] = 0$ w.p.1. However, in many practical cases, one uses a recursion like (4), but where the $\{\xi_n\}$ constitute a sequence of actual observation errors, which can not be represented in terms of the laws such as $H(\cdot|x)$ or $H'(\cdot|x)$. Various examples appear in Ljung [12, 13], and the references contained therein. Here, we work with (4), and no characterization of $\{\xi_n\}$ is involved.

Discussion of condition (A5). Suppose that there is a real sequence $\{\sigma_n^2\}$ such that the $\{\xi_n\}$ satisfy the usually imposed condition in stochastic approximation

$$
\begin{aligned}
&E[\xi_{n+1} | \ \xi_i,\ i \le n;\ a_i \ \ i \le n+1;\ X_0] = 0 \quad \text{w.p.1.} \\
&E[|\xi_{n+1}|^2 \ | \xi_i,\ i \le n;\ a_i,\ i \le n+1;\ X_0] \le \sigma_{n+1}^2, \ \text{w.p.1.}
\end{aligned} \tag{5}
$$

Then to verify (3), we use the martingale property of the partial sums of the $a_i\xi_i$ and get

$$
E \max_{0\le t\le T} \left| \sum_{i=m(t_N)}^{m(t_N+t)-1} a_i\xi_i \right|^2 \le 4E \sum_{i=m(t_N)}^{m(t_N+T)-1} a_i^2\sigma_i^2
$$

$$
E \max_{0\le t\le T} \left| \sum_{i=m(t_N-T)}^{m(t_N-t)-1} a_i\xi_i \right|^2 \le 4E \sum_{i=m(t_N-T)}^{m(t_N)-1} a_i^2\sigma_i^2.
$$

Then (3) holds if

$$\sup_{n \geq i} a_n \sigma_n^2 \to 0, \text{ as } i \to \infty. \tag{6}$$

Condition (6) holds if $\{\sigma_n^2\}$ is uniformly bounded (e.g., if $H'(\cdot|x)$ is used). If $H(\cdot|x)$ is used, then σ_n^2 is inversely proportional to c_n^2. Suppose that there are real $A, C, \gamma, \alpha, \sigma^2$ such that $a_n \leq An^{-\alpha}, \sigma_n^2 = \sigma^2 c_n^{-2} \leq \sigma^2 C^2 n^{-2\gamma}$ and $\alpha > 2\gamma$, then (6) also holds.

Now, let us drop (5). Suppose that the $\{a_i\}$ are real numbers. (If $\{a_i\}$ are random variables, then we can still continue the argument by use of suitable upper bounding of the $\{a_i\}$.) For notational convenience suppose that the ξ_i are scalar valued. (Otherwise we treat each component separately.) Define $E\xi_i\xi_j\xi_k\xi_\ell \equiv r_{ijk\ell}$. Suppose that there are functions $\bar{R}_{ij}$ such that $\bar{R}_{ij} = \bar{R}_{ji}$ and

$$|r_{ijk\ell}| \leq \bar{R}_{ij}\bar{R}_{k\ell} + \bar{R}_{ik}\bar{R}_{j\ell} + \bar{R}_{i\ell}\bar{R}_{kj}.$$

Condition (3) holds if there is a function $R(\cdot)$ such that $\bar{R}_{ij} \leq R(|i-j|)$ and $R(n) \to 0$ as $n \to \infty$. We omit the argument, for it can be seen from the following more general case, where the covariance matrix of ξ_i is inversely proportional to c_i^2. Assume that

$$\bar{R}_{ij} \leq R(|i-j|)/c_i c_i. \tag{7}$$

It is implied by Theorem 15.3, and the proof of Theorem 12.3 of Billingsley [1] that (3) will hold if there is a real sequence $\{K_N\}$ such that

$$(8)\qquad E\left|\sum_{m(t_N-s)}^{m(t_N+t)-1} a_i\xi_i\right|^4 \le K_N|t+s + a_{m(t_N-s)-1}|^2, \text{ all } t \ge 0, \quad t_N \ge s \ge 0,$$

where $K_N \to 0$ as $N \to \infty$. The left side of (8) is bounded above by

$$\sum_{i,j,k,\ell} a_i a_j a_k a_\ell (\bar{R}_{ij}\bar{R}_{k\ell} + \bar{R}_{ik}\bar{R}_{j\ell} + \bar{R}_{i\ell}\bar{R}_{kj}).$$

(The summation for each index is from $m(t_N-s)$ to $m(t_N+t)-1$. The limits will be omitted for notational simplicity.)

Consider one term only, and note that

$$(9)\qquad \sum_{i,j,k,\ell} a_i a_j a_k a_\ell \bar{R}_{ij}\bar{R}_{k\ell} \le \left[\sum_{i,j} a_i a_j \frac{R(|i-j|)}{c_i c_j}\right]^2 \le \left[2\sum_{\substack{i,j \\ i\ge j}} a_i a_j \frac{R(|i-j|)}{c_i c_j}\right]^2.$$

Suppose that

$$\limsup_{i\to\infty} \sum_{j\ge i} a_j/c_i c_j = 0.$$

and that $R(n)$ is summable. Then

$$(10)\qquad \sum_{j\ge i} a_j R(|i-j|)/c_i c_j \to 0$$

as $i \to \infty$. This implies that there is a K_N of the required type such that

$$\sum_{j,i=m(t_N-s)}^{m(t_N+t)-1} a_i a_j R(|i-j|)/c_i c_j \le (\tfrac{K_N}{3})^{1/2} \sum_{i=m(t_N-s)}^{m(t_N+t)-1} a_i,$$

which implies (8).

Define processes $X^o(t)$, $F^o(t)$, $B^o(t)$, $E^o(t)$ on $(-\infty,\infty)$ by $X^o(t) = X_0$, $B^o(t) = E^o(t) = F^o(t) = 0$ for $t \le 0$, and for $t \ge 0$,

$$X^o(t) = X_n \text{ on } [t_n,t_{n+1}),\ F^o(t) = \sum_{i=0}^{m(t)-1} a_i f_x(X_i),$$

$$B^o(t) = \sum_{i=0}^{m(t)-1} a_i \xi_i,\ E^o(t) = \sum_{i=0}^{m(t)-1} a_i \xi_i.$$

Then

(11) $$X^o(t) = X^o(0) - F^o(t) + B^o(t) + E^o(t), \quad t \in (-\infty,\infty).$$

Theorem 1. Assume (A1) to (A5), and that

(12) $$\sup_n |X_n| < \infty \text{ w.p.1.}$$

Then $\{X_n\}$ tend to S in probability, as $n \to \infty$. More strongly, for each $T < \infty$ and $\varepsilon > 0$,

(13) $$P\{\sup_{-T \le s \le T} \text{dist.}[X^o(t_N+s),S] \ge \varepsilon\} \to 0 \text{ as } N \to \infty.$$

(dist(x,S) denotes the Euclidean distance between x and the set S).

Remarks on the theorem. The condition (12) can often be verified by independent means. In practical problems some restriction would be put on the maximum values anyway. We get only convergence in distribution-or the somewhat stronger result (13), while the classical results concern convergence w.p.1. In fact, under our conditions, convergence w.p.1 is not always possible. But the conditions here are relatively weak.

Proof. By (11),

$$(14) \qquad X^o(t_N+s) = X^o(t_N) - [F^o(t_N+s) - F^o(t_N)] + [B^o(t_N+s) - B^o(t_N)] + E^o(t_N+s) - E^o(t_N)].$$

The idea of the proof exploits the fact that the convergence of $\{a_n\}$ to zero implies an increasing "compression of discrete time to continuous time" in the functions in (14), as $N \to \infty$. We expect that this "compression" and (A4) and (A5) will "eliminate", asymptotically, the effects of $B^o(\cdot)$ and $E^o(\cdot)$ relative to the effects of $D^o(\cdot)$. We define a sequence of left shifts so that the "asymptotic" part of the functions remain accessible for an analysis using weak convergence and differential equations methods. For each integer $N \geq 1$ define the processes $X^N(\cdot)$, $F^N(\cdot)$, $E^N(\cdot)$, $B^N(\cdot)$ on $(-\infty,\infty)$ by

$$
\begin{aligned}
&X^N(s) = X^o(t_N+s),\ X^N(0) = X^o(t_N) = X_N \\
&F^N(s) = F^o(t_N+s) - F^o(t_N) \\
&E^N(s) = E^o(t_N+s) - E^o(t_N) \\
&B^N(s) = B^o(t_N+s) - B^o(t_N).
\end{aligned} \tag{15}
$$

Then, (14) can be written as

$$
X^N(s) = X^N(0) - F^N(s) + B^N(s) + E^N(s). \tag{16}
$$

The fact that $E^N(s) \to 0$ in probability as $N \to 0$, for each s, follows from (A5). Similarly, $B^n(s) \to 0$ in probability as $N \to \infty$ for each s, by (A4) and the representation

$$
B^n(s) = -\sum_{m(t_N+s)}^{m(t_N)-1} a_i\beta_i,\ s < 0,\ B^N(s) = \sum_{m(t_N)}^{m(t_N+s)-1} a_i\beta_i,\ s > 0, \tag{17}
$$

and the fact that the sums of the a_i's over the above ranges is $\le [s+a_{m(t_N+s)-1}]$ $(s < 0)$ and $\le [s+a_{m(t_N)-1}] = [s+a_{N-1}]$ for $s > 0$.

The sequence $\{\Phi^N(\cdot),\ N = 1,2,\ldots\} = \{X^n(\cdot),F^N(\cdot),B^N(\cdot), E^N(\cdot),\ N = 1,2,\ldots\}$ has paths in $D^{4r}(-\infty,\infty)$, and we will now prove tightness and that the process which is the "limit" of any subsequence which converges in distribution has continuous paths w.p.1. We need only verify (1) and (2).

Criterion (1), (2) hold for $\{F^N(\cdot)\}$ by (12), the continuity of $f_x(\cdot)$, and the fact that $F^N(0) \equiv 0$. The criterion holds for $\{B^N(\cdot)\}$, by (A4) and the fact that $B^N(0) \equiv 0$. Condition (A5) and the fact that $E^N(0) = 0$, imply (1), (2) for $\{E^N(\cdot)\}$. Finally $\{X^N(0)\} = \{X_N\}$ satisfies (1), by (12). This last fact, together with the fact that the $\{F^N(\cdot)\}$, $\{B^N(\cdot)\}$ and $\{E^N(\cdot)\}$ satisfy (2) for each $T < \infty$, implies that $\{X^N(\cdot)\}$ satisfies (2) for each $T < \infty$. Thus $\{\Phi^N(\cdot)\}$ is tight on $D^{4r}(-\infty,\infty)$, and any process which is the limit in distribution of a convergent subsequence has continuous paths w.p.1.

Let N index a subsequence of $\{\Phi^N(\cdot)\}$ which converges in distribution, and let $\Phi(\cdot) = (X(\cdot),F(\cdot),B(\cdot),E(\cdot))$ denote the "limit". Then $B(t) \equiv E(t) \equiv 0$. Since

$$(18) \quad X^N(t) = X^N(0) - \int_0^t f(X^N(s))ds + \text{functions tending to the zero function in distribution,}$$

we must have

$$(19) \quad X(t) = X(0) - F(t) = X(0) - \int_0^t f_x(X(s))ds, \quad t \in (-\infty,\infty)$$

Since

$$\lim_{M\to\infty} \sup_N P\{\sup_{\infty>t>-\infty} |X^N(t)| \geq M\} = \lim_{M\to\infty} P\{\sup_{\infty>n\geq 0} |X_n| \geq M\} = 0,$$

the weak convergence implies that

$$P\{\sup_{\infty>t>-\infty} |X(t)| < \infty\} = 1.$$

The only solutions to (19) which are uniformly bounded on $(-\infty,\infty)$ must be in the largest finite invariant set of $\dot{x} = -f_x(x)$. But, since $f_x(\cdot)$ is the gradient of a function that is bounded from below, this invariant set is just S. Thus $X(0) \in S$. Hence $X^N(0) = X_N \to S$ in distribution.

Let T denote a positive real number. Since the function $q(\cdot)$ on $D^r(-\infty,\infty)$ which is defined by

$$q(y(\cdot)) = \min[1, \sup_{-T\leq s\leq T} \text{distance}(y(s),S)]$$

is continuous on $D^r(-\infty,\infty)$ w.p.1, relative to the measure induced on $D^r(-\infty,\infty)$ by the continuous process $X(\cdot)$, the weak convergence implies that the distribution of $q(X^N(\cdot))$ must converge to that of $q(X(\cdot))$. Clearly, the subsequence is irrelevant, since S does not depend on the subsequence and $q(X(\cdot)) = 0$ for any limit $X(\cdot)$. Thus (13) holds. Q.E.D.

3. Extensions. Several interesting extensions will be indicated. They are suggestive of additional applications. The details are similar to those concerned with Theorem 1, and only sketches will be given

(a) Continuous parameter stochastic approximations. Let $a(\cdot)$ denote a real valued positive measurable function on $[0,\infty)$, and suppose that $a(t) \to 0$ as $t \to \infty$ and $\int_0^\infty a(s)ds = \infty$. Assume (A2) and (A3), and let $\beta(\cdot)$ denote a R^r valued measurable

process on $[0,\infty)$ which tends to zero as $t \to \infty$. Let $\xi(\cdot)$ denote a measurable R^r valued process on $[0,\infty)$. Define the stochastic approximation (20)

$$(20) \qquad X(t) = X(0) - \int_0^t a(s) f_x(X(s))ds + \int_0^t a(s)\beta(s)ds + \int_0^t a(s)\xi(s)ds.$$

Such continuous time procedures have been discussed by Sakrison [14] and others.

Define the time transformation $\tau(\cdot)$ by

$$t = \int_0^\tau a(s)ds,$$

and define the function $Y^o(\cdot)$ by $Y^o(t) = X(\tau(t))$. Then

$$Y^o(t) = Y^o(0) - \int_0^t f_x(Y^o(s))ds + \int_0^{\tau(t)} a(s)\beta(s)ds$$

$$+ \int_0^{\tau(t)} a(s)\xi(s)ds.$$

For each real $T \geq 0$ define the functions $Y^T(\cdot), F^T(\cdot), E^T(\cdot)$ and $B^T(\cdot)$ on $(-\infty,\infty)$ by

$$Y^T(t) = Y^o(T+t),$$

$$F^T(t) = \int_T^{T+t} f_x(Y^o(s)ds, \quad B^T(t) = \int_{\tau(T)}^{\tau(T+t)} a(s)\beta(s)ds$$

$$E^T(t) = \int_{\tau(T)}^{\tau(T+t)} a(s)\,\xi(s)ds, \qquad \infty > t \geq -T.$$

At $t \leq -T$, let the functions take their values at $-T$. Then

$$Y^T(t) = Y^T(0) - F^T(t) + B^T(t) + E^T(t), \tag{21}$$

which is just the continuous parameter version of (16). Assume that

$$\sup_{t\geq 0}|X(t)| < \infty \quad \text{w.p.1.} \tag{22}$$

The sequence $\{Y^T(\cdot), F^T(\cdot), B^T(\cdot), E^T(\cdot),\ T < \infty\}$ has its paths in $C^{4r}(-\infty,\infty)$, the space of R^r valued continuous functions on $(-\infty,\infty)$. We can, of course, assume that they are in $D^{4r}(-\infty,\infty)$ and verify criterion (1)-(2) on each interval $[-T,T]$.

The sequence $\{B^T(\cdot)\}$ is tight and asymptotically degenerate. Also $\{F^T(\cdot)\}$ is tight, by (22) and the continuity of $f_x(\cdot)$. If $\{E^T(\cdot)\}$ were tight and asymptotically degenerate, then arguments parallel to those used in the proof of Theorem 1 would yield that if $Y(\cdot)$ is any limit in distribution of $\{Y^N(\cdot)\}$, then $Y(\cdot)$ must satisfy

$$\dot{Y}(t) = -f_x(Y(t)),$$

and $Y(t) \in S$, all $t \in (-\infty,\infty)$, and also that

$$X(t) \to S,\ t \to \infty, \quad \text{and}$$

$$P\{\sup_{-t\leq s\leq t} \text{dist}(X(T+s),S) \geq \varepsilon\} \to 0,\ T \to \infty, \text{ for any } t > 0.$$

Tightness of $\{E^T(\cdot)\}$ is implied by (1) and (2) (for $E^T(\cdot)$ replacing $X_n(\cdot)$), which, in turn, is implied by a continuous parameter analog of (3), where the sum is replaced by an integral, t_N by T and $m(u)$ by $\tau(u)$, for each $u \in (-\infty,\infty)$. For example, suppose (for notational convenience only) that $\xi(\cdot)$ is scalar valued and that

$$|E\xi(v_1)\xi(v_2)\xi(v_3)\xi(v_4)| \leq r(v_1,v_2,v_3,v_4),$$

where $r(\cdot)$ satisfies

$$r(v_1,v_2,v_3,v_4) \leq \bar{R}(v_1,v_2)\bar{R}(v_3,v_4) + \bar{R}(v_1,v_3)\bar{R}(v_2,v_4) + \bar{R}(v_1,v_4)\bar{R}(v_2,v_3),$$

and that there are positive measurable functions $c(\cdot)$ (a finite difference interval) and $R(\cdot)$ such that

$$\bar{R}(v_1,v_2) \leq R(|v_1-v_2|)/[c(v_1)c(v_2)]$$

Suppose that there is a sequence $K_T \to 0$ as $T \to \infty$ such that for each t,s with $t > s$,

$$(23) \quad \int_{\tau(T+s)}^{\tau(T+t)} \int \frac{a(v)a(w)}{c(v)c(w)} R(|v-w|)dvdw \leq K_T^{1/2}|t-s|,$$

Then we can show that there is a real number K such that

$$E|E^T(t) - E^T(s)|^4 \leq K \cdot K_T|t-s|^2,$$

which implies both tightness and asymptotic degeneracy of

$\{E^N(\cdot)\}$. Equation (23) does not seem to be particularly stringent in applications.

(b) A nondegenerate stochastic approximation. Another interesting application of the ideas of the last section occurs if we wish to design a procedure to find an actual absolute minimum of $f(\cdot)$, rather than simply a stationary point. In order to prevent the stochastic approximation procedure from degenerating to a stationary point, and to continue the search until "all" local minima are found, we may force the process to jump to a new "starting position" every once in a while. This is not necessarily the best procedure, but is easy to implement and often discussed. Suppose that $g(\cdot)$ is a bounded continuous function on R^r which is zero for large x, let $b > 0$ be a real number and let $Q(\cdot)$ denote a distribution function on R^r, with support in a bounded set. Let j_n denote a sequence of R^r valued random variables, such that $P\{j_n = 0 | a_i, i \leq n; j_i, i < n\} = 1 - a_n b + o(a_n)$ and $P\{j_n \in A | j_n \neq 0; j_i, i \leq n; a_i, i \leq n\} = Q(A)$. Define $J^0(t) = \sum_{i=0}^{m(t)-1} j_i$, $J^N(t) = J(t+t_N) - J(t_N)$, and the iteration

$$X_{n+1} = X_n - a_n f_x(X_n) + a_n B_n + a_n \xi_n + g(X_n) j_n. \tag{24}$$

Let $J(\cdot)$ denote a Poisson jump process with infinitesimal jump probability bdt, and jump distribution $Q(\cdot)$, and defined on $(-\infty,\infty)$. Under (A1) - (A5) and (12), a generalization of the method of proof of Theorem 1 yields that $X^N(\cdot)$ tends in distribution to a process $X(\cdot)$ which solves

$$(25) \qquad X(t) = X(0) - \int_0^t f_x(X(t))ds + \int_0^t g(X(s^-))dJ(s), \quad t \in (-\infty,\infty),$$

and which is uniformly bounded (pathwise) on $(-\infty,\infty)$. In (24), or (25), each time there is a jump, the search is "renewed", in a sense.

(c) Constrained stochastic approximations. Classical stochastic approximation basically consists of sequential Monte-Carlo procedures for finding zeroes or minima of regression functions. Many problems in applications involve constraints, although much less work has been done on the constrained problem (see Fabian [4], Kushner [7], Kushner and Sanvicente [8]). We will discuss a version of Theorem 1 for a simple problem with equality constraints.

Let $\phi(\cdot)$ denote a continuously differentiable R^q valued function on R^r, $q < r$ and let $\Phi(x)$ denote the Jacobian matrix of $\phi(x)$. We wish to find a parameter x which minimizes $f(\cdot)$ over the feasible set $B = \{x: \phi(x) = 0\}$. The function $f(\cdot)$ is not known, but, at any parameter setting X_n, we can observe (as for the classical stochastic approximation) a random variable which we write as $f_x(X_n) + a_n\beta_n + a_n\xi_n$, where β_n and ξ_n represent an observation "bias" and observation "noise", resp.

Assume that $\Phi'(x)\phi(x) = 0$ implies that $\Phi(x)$ is of full rank (hence that $\phi(x) = 0$ also). Some such assumption is usually required of numerical algorithms, to assure that the iterates will not "get stuck" outside of B, even in the purely deterministic case.

In the unconstrained case, the algorithm (4) seeks a stationary point of $f(\cdot)$; i.e., a point x where the necessary condition for minimality $f_x(x) = 0$ holds. In the constrained case, we also seek a feasible point where a necessary condition for optimality holds (the usual one of the calculus). In particular, we seek points $x \in B$ such that there is a vector λ of Lagrange multipliers, with components $\lambda_1,\ldots,\lambda_q$ such that

$$f_x(x) + \Phi'(x)\lambda = 0. \tag{26}$$

Let S denote the set of such feasible points and suppose that $S \cap B$ is bounded, and connected.

Let k denote a fixed positive number. The algorithm (27) was discussed by Kushner and Kelmanson [9], who proved w.p.1 convergence to the set $S \cap B$ under the conditions (among others) that ξ_i satisfied (5) and that the $\{a_i\}$ were square summable.

$$X_{n+1} = X_n - a_n[\pi_n(f_x(X_n)) + \pi_n\xi_n + \pi_n\beta_n + k\Phi'(X_n)\phi(X_n)], \tag{27}$$

where we use the definitions

$$\pi(x) = [I - \Phi'(x)(\Phi(x)\Phi'(x))^{-1}\Phi(x)]$$

and

$$\pi_n = \pi(X_n).$$

The inverse notation implies that the pseudo inverse is used if the inverse doesn't exist. The matrix $\pi(x)$ is just a projection onto the orthogonal complement of the rows of $\Phi(x)$. If x is feasible and $\pi(x)f_x(x) = 0$, then $x \in S \cap B$.

Assume (A1) - (A5) and (12). Define the functions $X^N(\cdot)$, $N \geq 0$, for the sequence generated by (27), as they were defined in Section 1 for the sequence generated by (4). Then $\{X^N(\cdot)\}$ is tight on $D^r(-\infty,\infty)$, and if N indexes a subsequence which converges in distribution, then the limit $X(\cdot)$ satisfies

$$\dot{X}(t) = -\pi(x)f_x(x) - k\Phi'(x)\phi(x) \quad \text{on} \quad (-\infty,\infty). \tag{28}$$

Next, we check that $X(t) \in B$. Consider the Liapunov function $P(x) = \phi'(x)\phi(x)$. Computing the derivative $\dot{P}(X(t))$ along trajectories, we get (at $X(t) = x$)

$$\dot{P}(x) = -2\phi'(x)\Phi(x)[\pi(x)f_x(x) + k\Phi'(x)\phi(x)].$$

Since $\pi(x)$ projects onto the subspace orthogonal to that spanned by the rows of $\Phi(x)$, $\Phi(x)\pi(x)\cdot v = 0$ for any vector v. Hence,

$$\dot{P}(x) = -2|\phi'(x)\Phi(x)|^2. \tag{29}$$

Since $\Phi'(x)\phi(x) = 0 \Longrightarrow \phi(x) = 0$, (29) and the fact that $P(x) \geq 0$, imply that the only invariant sets M for the system (29) must satisfy the property: if $x \in M$, then $\phi(x) = 0$. Next, using the fact that $\pi(x)$ is a projection, and $f_x(\cdot)$ is a gradient of a function that is bounded from below, we get that the invariant sets must be included in $\{x: \pi(x)f_x(x)\} = 0$. Thus the invariant sets are in $B \cap S$. Following the reasoning in the conclusion of the proof of Theorem 1, we see that $X_n \to B \cap S$ in distribution as $n \to \infty$, and (13) holds.

(d) The technique can clearly be applied to Robbins-Munro procedures, and to other algorithms of the form (4), where $f_x(\cdot)$ is replaced by a suitable alternative.

The technique is useful for getting rates of convergence. This topic will be developed in a subsequent paper, for various constrained and unconstrained problems. Even, the conditions on $f_x(\cdot)$ can be weakened, provided that a proper flow can be defined for $x = f_x(\cdot)$.

REFERENCES

[1] Billingsley, P.; Convergence of Probability Measures, Wiley, New York, 1968.

[2] Blum, J.; "Multidimensional stochastic approximation methods", Ann. Math. Statist., 25, 1954, pp. 734-744.

[3] Driml, M., Nedoma, J.; Stochastic Approximation for Continuous Random Processes', Proc. 2nd Prague Conf. on Information Theory and Statistical Decision Functions, Prague, 1959.

[4] Fabian, V.; "Stochastic approximation of constained minima", Proc. 4th Prague Conference on Statistical Decision Theory and Information theory, 1966, pp. 277-289.

[5] Fabian, V., "Stochastic Approximation", Optimiz. Methods in Statistics, 1971, pp. 439-470, ed. by Rustagi, Academic Press, New York.

[6] Kiefer, J., Wolfowitz, J.; "Stochastic estimation of the maximum of a regression function", Ann. Math. Statist., 23, 1952, pp. 462-466.

[7] Kushner, H.J.; "Stochastic approximation algorithms for constrained optimization problems", Ann. Statist., 2, 1974, pp. 713-723.

[8] Kushner, H.J., Sanvicente, E.; "Stochastic approximation for constrained systems with observation noise on the systems and constraints", Automatica, 11, 1975, pp. 375-380.

[9] Kushner, H.J., Kelmanson, M.L.; "Stochastic approximation algorithms of the multiplier type for the sequential Monte Carlo optimization of stochastic systems", to appear in SIAM J. on Control, 1976.

[10] LaSalle, J.P., Lefschetz, S.; Stability by Liapunov's Direct Method, Academic Press, New York, 1961.

[11] Lindvall, T.; "Weak convergence of probability measures and random functions in the function space $D[0,\infty)$", J. Appl. Prob., 10, 1973, pp. 109-121.

[12] Ljung, L.; "Convergence of recursive stochastic algorithms", report 7403, 1974, Lund Institute of Technology, Division of Automatic Control, Lund, Sweden.

[13] Ljung, L., Soderstrom, T., Gustavsson, I.; "Counterexamples to general convergence of a commonly used recursive identification method", IEEE Trans. on Automatic Control, AC-20, 1975, pp. 643-652.

[14] Sakrison, D.J.; "A continuous Kiefer-Wolfowitz procedure for random process", Ann. Math. Statist., 35, 1964, pp. 590-599.

[15] Schmetterer, L.; "Stochastic approximation", pp. 587-608, 4th Berkeley Symposium on Probability and Statistics, 1961, Univ. of California Press, Berkeley.

[16] Wasan, M.T., Stochastic Approximation, Cambridge University Press, 1969.

II. RATES OF CONVERGENCE FOR SEQUENTIAL MONTE CARLO OPTIMIZATION METHODS[+]

by

Harold J. Kushner
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, R. I. 02912

August 4, 1976

[+]This research was supported in part by the Office of Naval Research under N000-14-76-C-0279, in part by the National Science Foundation under 73-03846-A01 and in part by the Air Force Office of Scientific Research under AFOSR 76-3063.

RATES OF CONVERGENCE FOR SEQUENTIAL MONTE CARLO OPTIMIZATION METHODS

Harold J. Kushner

Abstract

Sequential Monte Carlo Methods of the Stochastic Approximation (SA) type, with and without constraints, are discussed. The rates of convergence are derived, and the quantities upon which the rates depend, are discussed. Let $\{X_n\}$ denote the SA sequence and define $U_n = (n+1)^{\beta}X_n$ for a suitable $\beta > 0$. The $\{U_n\}$ are interpolated into a natural continuous time process, and weak convergence theory is applied to develop the properties of the tails of the sequence. The technique has a number of advantages over past approaches - advantages which are discussed in the paper. It gives more insight (and is apparently more readily generalizable) than do other approaches - and suggests ways of improving the convergence. The particular "dynamical" nature of the approach allows one to say more about the "tail" process - and to do more "decision" (or "control") analysis with it.

RATES OF CONVERGENCE FOR SEQUENTIAL MONTE CARLO OPTIMIZATION METHODS

Harold J. Kushner

1. Introduction.

The subject of stochastic approximation (SA) for unconstrained systems has been well developed in many respects over the past 25 years. See, e.g., the references in Wasan [1], and also Ljung [2,3]. The treatment of SA under constraints is relatively recent; see Fabian [4], Kushner [6], Kushner and Gavin [5], Kushner and Sanvicente [7], [8], Kushner and Kelmanson [9], and Kushner [10]. The SA problem (with or without constraints) occurs when wishes to choose a parameter $x \in R^r$ (Euclidean r-space) of a system which (at least locally) minimizes a scalar valued performance function $f(x)$ (without or under constraints), but where the form of $f(\cdot)$ is unknown (as it usually is in complex control problems), and where only noise corrupted measurements of the performance can be made, at various selected parameter settings. The algorithms give a sequence of parameter values $\{X_n\}$ which converges to a local minimum in some statistical sense. The subject is a stochastic Monte-Carlo form of the general computational problem of non-linear programming, and has numerous applications to control theory and practice in the areas of optimization identification and tracking.

All of the previous works on the constrained problem treat only the fact of convergence. Here we give results on rates of

convergence, and obtain some new results for the unconstrained problem also. In particular, we will show that when suitably scaled and interpolated, the "tail" of the SA process converges weakly to a linear diffusion process. The scaling and properties of the diffusion give the rates of convergence, and much interesting additional information as well. Some of the advantages will be made clear below.

In Section 2, the unconstrained algorithm is introduced. Sections 3 and 4 introduce a Lagrangian and augmented penalty function algorithms. The unconstrained problem is further developed in Section 5, and the theorem stated. Section 6 contains some background on weak convergence of a sequence of probability measures on certain metric spaces. This theory is a very natural tool for analyzing the asymptotic properties of the interpolation, and for obtaining useful information on this "tail". In particular, it enables us to exploit the statistical structure of the tail, to enhance the rate of convergence. The method of proof is new in SA, and seems to be more easily generalizable (to other types of noise sequences and to other types of constrained problems) than past approaches. Relatively little is known concerning the statistical properties of SA sequences, considered as a process. Our approach seems to be a useful tool for dealing with such properties, for it emphasizes the "process" aspects of the problem. See, also the remarks after the statement of Theorem 5.1.

2. The Unconstrained Case. Formulation.

Let X_n (with components $X_n^1,\ldots,X_N^r$) denote the n^{th}

estimate of the local minimum, and let e_i denote the unit vector in the i^{th} coordinate direction, and let $\{a_n, c_n\}$ be null sequences, a_n being a positive definite matrix, and c_n a (finite difference interval) positive scalar. Define the "observation difference" $\delta Y_n = \{\delta Y_n^1, \ldots, \delta Y_n^r\}$ and the <u>observation noises</u> $\xi_n^{i,1}, \xi_n^{i,2}$ by

$$\begin{aligned}\delta Y_n^i &= (\text{observation at parameter } (X_n + e_i c_n)) - \\ &\quad (\text{observation at parameter } (X_n - e_i c_n)) \\ &\equiv [f(X_n + e_i c_n) + \xi_n^{i,1}] - [f(X_n - e_i c_n) + \xi_n^{i,2}]\end{aligned}$$

Let $\xi_n^i = \xi_n^{i,1} - \xi_n^{i,2}$ and $\xi_n = (\xi_n^1, \ldots, \xi_n^r)$, and let $\mathcal{B}_n$ denote the σ-algebra determined by $X_0, \ldots, X_n, \xi_0, \ldots, \xi_{n-1}$.

Define X_{n+1} by $(\delta f_n \equiv (\delta f_n^1, \ldots, \delta f_n^r)$, $\delta f_n^i \equiv f(X_n + e_i c_n) - f(X_n - e_i c_n))$

$$X_{n+1} = X_n - \frac{a_n}{2c_n}\delta Y_n = X_n - a_n\{\frac{\delta f_n}{2c_n} + \frac{\xi_n}{2c_n}\}. \tag{2.1}$$

The w.p.1, convergence of $\{X_n\}$ (when there is only one stationary point of $f(\cdot)$) has been the subject of most of the references in Wasan [1], and we mention only some of the conditions usually assumed, namely:

$$E_{\mathcal{B}_n}\xi_n = 0 \text{ w.p.1}, \quad E_{\mathcal{B}_n}\xi_n\xi_n' \leq M, \text{ all } n, \text{ for some matrix } M. \tag{2.2}$$

(2.3) $$\sum_n a_n = \infty$$

(2.4) $$\sum_n a_n c_n < \infty, \quad \sum_n |a_n|^2 / c_n^2 < \infty.$$

The conditions were considerably relaxed in Ljung [2] and in Kushner [10], although they required more smoothness on $f(\cdot)$ than the previous works did. Also, [10] proved convergence to a stationary point of $f(\cdot)$, even when not unique. The conditions required here will be given later.

By simple alterations in the calculations, it is possible to treat non-central difference and continuous time forms.

3. The Constrained Problem. A Lagrangian Method.

Now, suppose that we wish to modify the problem so that $f(x)$ is minimized under constraints $q_i(x) \leq 0$, $i = 1,\ldots,s$, where each $q_i(\cdot)$ is continuously differentiable. Let $Q(x)$ denote the matrix $\{q_{1,x}(x),\ldots,q_{s,x}(x)\}$, where $q_{i,x}(\cdot)$ is the gradient of $q_i(\cdot)$, and let $\{b_n^i\}$, $i = 1,\ldots,s$, denote positive null sequences and let $\lambda = (\lambda^1,\ldots,\lambda^s)$, $\lambda^i \geq 0$. Consider the Lagrangian algorithm.

(3.1) $$\lambda_{n+1}^i = \max[0, \lambda_n^i + b_n^i q_i(X_n)], \quad i = 1,\ldots,s,$$

(3.2) $$X_{n+1} = X_n - a_n\{\frac{\delta Y_n}{2c_n} + Q(X_n)\lambda_n\}.$$

Suppose that there is a known number M such that the constrained minimum θ, and the corresponding multipliers $\overline{\lambda}$ satisfy $|\theta^i| \leq M$, $\overline{\lambda}^i \leq M$. Modify (3.1), (3.2) by projecting on to the sets $\{x: |x^i| \leq M\}$, $\{\lambda: \lambda^i \leq M\}$, whenever the bounds are exceeded. Then, under conditions (2.2) - (2.4), and convexity conditions on $f(\cdot)$ and $q(\cdot)$, [7] proved that $\{X_n\}$ converges w.p.1. to the constrained minimum θ.

It was not proved that λ_n converged to an optimal multiplier. Indeed, the optimal multiplier may not be unique. Yet, let us note for later use, that in very many of the examples which we simulated, it appeared that λ_n did converge to a $\overline{\lambda}$ such that the Kuhn-Tucker condition

$$f_x(\theta) + Q(\theta)\overline{\lambda} = 0 \tag{3.3}$$

held.

4. Equality Constraints. An Augmented Penalty Function Method.

Suppose now that we wish to minimize $f(\cdot)$ subject to equality constraints $\phi_i(x) = 0$, $i = 1,\ldots,s$, where $\phi_i(\cdot)$ are continuously differentiable. An SA version of Mieles' [11] augmented penalty function method was developed in [9]. Let $0 < k$ denote a real number, define $P(x) = \frac{1}{2}\sum_i |\phi_i(x)|^2$ and $\Phi(x) = \{\phi_{1,x}(x),\ldots,\phi_{s,x}(\cdot)\}'$. Let $\pi(x)$ denote the operator:

$(I-\pi(x))v$ is the projection of $v \in R^r$ onto the span of $\{\phi_{1,x}(x),\ldots,\phi_{s,x}(\cdot)\}$. The necessary condition of the calculus for a local stationary point at x is $\pi(x)f_x(x) = 0$.

Define the algorithm $(P_x(x) = \Phi'(x)\phi(x))$

$$(4.1) \qquad X_{n+1} = X_n - a_n[\pi(X_n)\frac{\delta Y_n}{2c_n} + k\Phi'(X_n)\phi(X_n)].$$

Under essentially the conditions (2.2) - (2.4), bounded double differentiability of $f(\cdot)$, and that $\Phi'(x)\phi(x) = 0$ implies that $\Phi(x)$ is of full rank, reference [9] proved convergence w.p.1 to a θ such that $\pi(\theta)f_x(\theta) = 0$.

5. Unconstrained Problem. Theorem Statement.

Return to the algorithm of Section 2. Let $a_n = A/(n+1)^\alpha$, $c_n = C/(n+1)^\gamma$, where C,α and γ are positive real numbers, $\alpha > \gamma$, and A is a positive definite matrix. With a bit of extra complication in the notation, rather general a_n,c_n sequences (which do not decrease too fast) can be handled. Let us list the following assumptions.

(A5.1) $f(\cdot)$ is continuous, and has bounded and continuous mixed second derivatives.

(A5.2) $\sum_n a_n^2 < \infty$.

(A5.3) $\sum_n a_n = \infty$.

(A5.4)[+] There is $\theta \in R^r$ such that $X_n \to \theta$ w.p.1.

(A5.5) $f(\cdot)$ has continuous third derivatives $f_{x_i x_i x_i}(x)$ at $x = \theta$. Define $B(\theta)$ = vector whose i^{th} component is this third derivative divided by 3!

(A5.6) $E_{\mathcal{B}_n} \xi_n = 0$ w.p.1.

(A5.7) There is a matrix $\Sigma(\theta)$ such that $E_{\mathcal{B}_n} \xi_n \xi_n' \to \Sigma(\theta)$ w.p.1, as $n \to \infty$.

(A5.8) For some $\delta > 0$, $M_1 < \infty$,

$$E_{\mathcal{B}_n} |\xi_n|^{2+\delta} \leq M_1 \quad \text{w.p.1, all } n.$$

(A5.9I) Let $\alpha = 1$, $\beta = \alpha/3 = 2\gamma$. Define $F(\theta)$ = Jacobian matrix of $f(\cdot)$ at θ. Let the eigenvalues of $AF(\theta) - \beta I \equiv \overline{K}_1$ have positive real parts.

or

[+]It is possible to treat the case where θ is a random variable. (A5.9) limits our consideration to rates for the sequences which converge to a strict local minimum.

(A5.9II) Let $\alpha < 1$, $\beta = 2\gamma = \alpha/3$, and let the eigenvalues of $AF(\theta) \equiv \bar{K}_2$ have positive real parts.

(A5.10) $f_x(\theta) = 0$.

Remark. The conditions are mostly self-explanatory. (A5.9) implies that θ is a strict local minimum. The w.p.1 convergence (A5.4) is assumed, because we are concerned with rates of convergence. The actual convergence is proved in the various cited references. Condition (A5.6) is essentially a classical condition in the subject. As discussed in Section 11, the condition can be readily weakened provided that we can still show that $P\{|U_n| \geq N\} \to 0$ as $N \to \infty$, uniformly in n, where $\{U_n\}$ is defined below. Indeed, the possibility of such extensions is one of the advantages of our approach.

Let $\varepsilon_{i,n}$ and $\bar{\varepsilon}_{i,n}$ denote functions whose values may differ from usage to usage, but which depend on X_n and c_n, and tend to zero w.p.1, as $n \to \infty$. (In Section 9, they may also depend on λ_n.) Define $\delta X_n = X_n - \theta$. Then, using (A5.4,5,10), (2.1) can be rewritten in the form

$$X_{n+1} = X_n - a_n[F(\theta)\delta X_n + B(\theta)c_n^2 + \varepsilon_{1,n}c_n^2 + \varepsilon_{2,n}\delta X_n] - a_n\xi_n/2c_n. \tag{5.1}$$

The next step is to scale $\{\delta X_n\}$. For some $\beta > 0$ (to be selected below - we are obviously interested in the largest β for

which the process $\{U_n\}$ makes sense) define $U_n = (n+1)^\beta \delta X_n$. Then, using $(n+2)^\beta = (n+1)^\beta(1+\beta/(n+1) + O(\frac{1}{n^2}))$,

$$(5.2) \qquad U_{n+1} = (I+\beta I/(n+1)-a_nF(\theta)-a_n\bar{\varepsilon}_{1,n})U_n$$
$$- a_n(n+1)^\beta c_n^2 B(\theta) - a_n(n+1)^\beta \xi_n/2c_n + a_n\bar{\varepsilon}_n,$$

where

$$\bar{\varepsilon}_n = (n+1)^\beta[\bar{\varepsilon}_{2,n}c_n^2 + \frac{1}{2c_n}\xi_n O(\frac{1}{n})].$$

It turns out that all limits of $\{U_n\}$ or of the interpolated $\{U_n\}$ introduced below do not depend on the (asymptotically negligable - for the β to be selected) $\{\bar{\varepsilon}_n\}$ sequence. To slightly simplify the development, we will drop the term henceforth, although its presence would not affect any of the subsequent arguments - except that an additional term would have to be carried.

Interpolation. Introduction. The next step in the formulation of the limit theorem involves an interpolation of $\{U_n\}$ into a continuous parameter process. The form of the interpolation is motivated by the following observation. Let $\{\psi_n\}$ denote a sequence of (zero mean) independent, identically distributed (for convenience here) random variables with unit variance and $E|\psi_n|^{2+\gamma} \leq M < \infty$ for some real $\gamma > 0$, $M > 0$, and D be a matrix whose eigenvalues have positive real parts. For each small $\Delta > 0$, define the sequence $\{V_n^\Delta\}$ and function $V^\Delta(\cdot)$ by

$$V^{\Delta}_{n+1} = (I-\Delta D)V^{\Delta}_n + \sqrt{\Delta}\ \psi_n, \quad V^{\Delta}_0 = x, \text{ fixed},$$

and $V^{\Delta}(t) = V^{\Delta}_n$ in $[n\Delta, n\Delta+\Delta)$. Then $\{V^{\Delta}(\cdot)\}$ converges in several statistical senses to the process solving

$$dV = -DVdt + dW,$$

where $W(\cdot)$ is a Wiener process.

Interpolation. Let $D[0,\infty)$ denote the space of real valued functions on $[0,\infty)$ which are right continuous and have left hand limits at each t. Suppose that $D[0,\infty)$ and its products are endowed with the Skorokhod topology (see Billingsley [16] for $D[0,T]$, Lindvall [20] for $D[0,\infty)$.) We mention only that convergence of a sequence $\{x^n(\cdot)\}$ to a continuous $x(\cdot)$ in that topology is equivalent to uniform convergence on each finite interval, and that, under that topology, the space is equivalent to a complete separable metric space, in that there is a metric, generating the same topology, under which the space is complete and separable (which we suppose henceforth).

Define $\Delta t_n = (n+1)^{-\alpha}$, $t_n = \sum_{i=0}^{n-1} \Delta t_i$, $t_0 = 0$, $\delta W_n = (n+1)^{\beta+\gamma-\alpha}\xi_n$ $= (n+1)^{\beta+\gamma-\alpha/2}(\xi_n\sqrt{\Delta t_n})$, $W_n = \sum_{i=0}^{n-1} \delta W_i$, $W_0 = 0$. For each integer n,N, define $W^N_n = W_{N+n} - W_N$, $\delta W^N_n = \delta W_{N+n}$, $U^N_0 = U_N$, and define $U^N(\cdot)$, $W^N(\cdot)$ by:

$$U^N(t) = U_{N+n},\ W^N(t) = W_{N+n} - W_N = W_n^N \quad \text{on} \quad [t_{N+n}-t_N, t_{N+n+1}-t_N).$$

Note that $a_n\xi_n(n+1)^{\beta}/2c_n = (A/2C)\delta W_n(1+n)^{\gamma+\beta-\alpha/2}$ and $a_n(n+1)^{\beta}c_n^2B(\theta) = (AB(\theta)C^2)\Delta t_n(n+1)^{\beta-2\gamma}$. Also, the paths of $W^N(\cdot)$ and $U^N(\cdot)$ are in $D^r[0,\infty)$.

Dropping the $a_n\bar{\varepsilon}_n$ term, we have

$$U_{n+1} = G_nU_n - (A/2C)\sqrt{\Delta t}_n\xi_n(n+1)^{\gamma+\beta-\alpha/2} - (AB(\theta)C^2)\Delta t_n(n+1)^{\beta-2\gamma}, \tag{5.3}$$

where

$$G_n = (I+\beta I/(n+1) - A\Delta t_n(F(\theta) + \bar{\varepsilon}_{1,n})).$$

It is clear from (5.3) that unless $\gamma + \beta - \alpha/2 \leq 0$, $\beta - 2\gamma \leq 0$, $E|U_n|^2$ will diverge. We use, henceforth, the maximum β, namely $\beta = 2\gamma = \alpha/3$, with which the exponents of $(n+1)$ in (5.3) are all zero.

Let $\bar{W}(\cdot)$ denote a standard R^r valued Wiener process and $\bar{U}(\cdot)$ the (stationary process) solution to

$$d\bar{U}(t) = -\bar{K}\bar{U}(t)dt - AB(\theta)C^2dt - (A/2C)\Sigma^{1/2}(\theta)d\bar{W}(t), \tag{5.4}$$

where $\bar{K} = \bar{K}_1$ or $\bar{K}_2$ (see A5.9).

The undefined terms (concerning weak convergence) in

Theorem 5.1 will be defined in the next section, and the proof given in Section 7.

Theorem 5.1. Under (A5.1) to (A5.10), $\{U^N(\cdot), W^N(\cdot)\}$ is tight on $D^{2r}[0,\infty)$, and $\{U^N(\cdot)\}$ converges weakly to the $\bar{U}(\cdot)$ of (5.4). (I.e., any weak limit has the probability law of $\bar{U}(\cdot)$ on $D^r[0,\infty)$ or on $C^r[0,\infty)$.)

Remarks. Clearly, we have the fastest rate of convergence when $\alpha = 1$ and A5.9I holds. The optimal normalization scaling β, and asymptotic normalized variance $(\lim_{t\to\infty} E\bar{U}(t)\bar{U}'(t))$ are not new; see Sacks [12] and Fabian [13], at least for this unconstrained problem. However, our framework is interesting for several reasons.

The technique emphasizes the behavior of the "tail" of $\{U_n\}$, considered as a dynamical process. The correlation structure of this process can sometimes be exploited to yield (at time N) a function of the $\{X_N, X_{N-1}, \ldots\}$, which is a better estimate of θ. See Section 8. In practice, we observe the path "dynamically", and it is worthwhile to try to understand its dynamical behavior. In certain cases, e.g., the Lagrangian method, the procedure often inherently oscillates around $\theta, \bar{\lambda}$, as it converges. The properties of this oscillation can be deduced from the relevant results of Section 9, and used to improve the estimate, or to design a more suitable process. The "constrained" results are new and rather interesting.

Perhaps the dynamical structure can lead to worthwhile

results on optimal stopping times - or to facilitate the use of SA in control processes, where the interest is often inherently dynamical. The proof has an independent interest. In particular, in the unnormalized case (which also used weak convergence methods and a "dynamical" approach) [10], we were able to prove convergence for a much wider and more realistic class of noise processes than those usually considered,and it is quite likely that the "rate" proofs can be extended to the wider types of noise sequences.(see Section 11 for a partial such result).

Extensions. By a slight change in the method of proof, we can also get the extensions:

I. Assume the conditions of Theorem 5.1, but set $\beta = \min[2\gamma,\alpha/3]$. If $2\gamma < \alpha/3$, (resp. $2\gamma > \alpha/3$) then noise (resp., bias) is relatively unimportant in the limit and the theorem holds with dw (resp., $B(\theta)$) set equal to zero. If $\alpha = 1$, and $0 < b < \beta$ and the eigenvalues of $[AF(\theta) - bI]$ have positive real parts, then the interpolation of $\{(n+1)^b(X_n-\theta)\}$ converges weakly to the zero process.

II. One-sided differences. Use the observation difference [observation at $(X_n+c_ne_i)$ - observation at $X_n]/c_n$ instead of δY_n^i. Then the theorem holds if $A/2C$, $B_i(\theta)$ and $\beta = 2\gamma = \alpha/3$ (or $\beta = \min[2\gamma,\alpha/3]$ in I above) are replaced by A/C, $f_{x_ix_i}(\theta)/2$ and $\beta = \gamma = \alpha/4$ (or $\beta = \min[\gamma,\alpha/4]$ in I above), resp.

Theorems 9.1 and 10.1 can also be extended in the same way.

6. Weak Convergence.

The material is in Billingsley [16]. See, also Whitt [17], Kushner [18, Chapter 2], Iglehart [19] or Lindvall [20], who gives the extensions of weak convergence on $D[0,T]$ (for some real T) to that on $D[0,\infty)$. Let $\{Z^n\}$ denote a sequence of random variables with values in a complete separable metric space S (such as $D[0,\infty)$, or $D^{r_1}[0,\infty) \times R^{r_2}$ or R^{r_2}, for some integers r_1, r_2), with induced measures $\{P^n\}$ on the Borel sets $\mathcal{S}$ of S. (On $D[0,\infty)$, $\mathcal{S}$ is also the Borel algebra over the coordinate projections; i.e. over the sets $\{x(\cdot): x(t) \in A = \text{real Borel}\}$.)

The $\{P^n\}$ sequence (and also $\{Z^n\}$, here) is said to be tight if for each $\varepsilon > 0$, there is a compact $K_\varepsilon \in \mathcal{S}$ such that $P^n(K_\varepsilon) \geq 1 - \varepsilon$, all n. If $\{P^n\}$ is tight, then any subsequence has a further subsequence which converges weakly to some measure P on $(S,\mathcal{S})$. If $\{P^n\}$ converges weakly to P, then

$$\int f(x)P^n(dx) \to \int f(x)P(dx)$$

for every bounded measurable $f(\cdot)$ which is continuous on a measurable set $S_0 \subset S$, such that $P(S_0) = 1$. If $(S,\mathcal{S}) = (S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$, and P_i^n is a measure on $(S_i, \mathcal{S}_i)$, $i = 1,2$, all n, then tightness of $\{P_1^n \times P_2^n\}$ on $(S,\mathcal{S})$ is implied by tightness of $\{P_i^n\}$ on $(S_i, \mathcal{S}_i)$, for each i.

If $P^n \to P$ weakly, and if Z is a S valued random variable with values in $(S,\mathcal{S})$ and with measure P, we say that $Z^n \to Z$ weakly also. In this sense, Theorem 5.1 is understood to mean that the measures of the $D^r[0,\infty)$ valued random variables $U^N(\cdot)$ (or the measures that $U^N(s)$, $s < \infty$ induce on $D^r[0,\infty)$) converge weakly to the measure that $\bar{U}(\cdot)$ induces on $D^r[0,\infty)$. In our case, Z^n will be identified with some combination of $U^n(\cdot), W^n(\cdot)$ and U_n. Also, it will be proved that $\{U^n(\cdot), W^n(\cdot), U_n\}$ is tight on $D^{2r}[0,\infty) \times R^r$.

Let $Z^n(\cdot)$ be a sequence of processes with paths in $D[0,\infty) = S$, w.p.1, and induced measures P^n on $(S,\mathcal{S})$. Then, there is tightness of $\{P^n\}$ or $\{Z^n(\cdot)\}$ on $D[0,\infty)$, if the restrictions to $D[0,T_k]$ are tight for some sequence $T_k \to \infty$. We sometimes use (without explicit mention) the fact (and similar facts) that if $E \max_{t \le T} |Z^n(t)| \to 0$ as $n \to \infty$ for each T, and $Z^n(\cdot) \in D[0,\infty)$ w.p.1, then $Z^n(\cdot)$ converges weakly to the zero element of $D[0,\infty)$. We note for future use, that if $h(\cdot,\cdot)$ is a bounded continuous function, and $Z^n(\cdot) \to Z(\cdot)$ weakly, where $Z(\cdot)$ has continuous paths, then the processes defined by $\int_0^t h(t,s)Z^n(s)ds$ converge weakly

to the process defined by $\int_0^t h(t,s)Z(s)ds$.

7. Proof of Theorem 5.1.

1^o. Returning to (5.3) and using $\beta = 2\gamma = \alpha/3$, we first show that $\{U_n\}$ is tight on R^r. Define (neglecting $a_n\bar{\varepsilon}_n$, as we can easily show is legitimite) $\{v_n\}$ and $\tilde{U}_n$ by

$$v_{n+1} = G_n v_n - AB(\theta)C^2\Delta t_n,$$

$$\tilde{U}_n = U_n - v_n.$$

Note that, if random functions Z_n and Z_n^ε take values in the same space for each $\varepsilon > 0$, and differ on at most an ω set of measure ε, and if $\{Z_n^\varepsilon\}$ is tight on some space for each $\varepsilon > 0$, then $\{Z_n\}$ is tight. Thus, since $\bar{\varepsilon}_{1,n} \to 0$ w.p.1, if tightness (and the theorem) is proved under the assumption that for each $\varepsilon > 0$, there is an integer $N_\varepsilon < \infty$ such that $|\bar{\varepsilon}_{1,n}| \leq \varepsilon$ for $n \geq N_\varepsilon$, and $|\bar{\varepsilon}_{1,n}|$ is uniformly bounded, then they will be true in general. We make the assumption on $\bar{\varepsilon}_{1,n}$. Under this assumption $\{v_n\}$ is bounded, hence tight on R^r. We only need to prove that $\{\tilde{U}_n\}$ is tight. We have

$$\tilde{U}_{n+1} = G_n\tilde{U}_n - (A/2C)\sqrt{\Delta t_n}\,\xi_n.$$

By (A5.6), (A5.8), and under cases (A5.9I) or (A5.9II), there are positive constants M_1, M_2 such that, for large n,

$$E_{\mathcal{G}_n}|\tilde{U}_{n+1}|^2 \leq |G_n|^2|\tilde{U}_n|^2 + M_1\Delta t_n$$

$$\leq (1-M_2\Delta t_n)|\tilde{U}_n|^2 + M_1\Delta t_n,$$

from which boundedness of $\{E|\tilde{U}_n|^2\}$, hence tightness of $\{\tilde{U}_n\}$ follows.[+]

2^o. <u>A representation for</u> $U^N(\cdot)$. Define $C_i^j = I$ for $i > j$ and $C_i^j = G_j \ldots G_i$ for $i \leq j$. Then (5.3) is solved to get

$$U_{N+n+1} = C_N^{N+n}U_n - \sum_{m=N}^{N+n} C_{m+1}^{N+n}[(A/2C)\delta W_m + AB(\theta)C^2\Delta t_m],$$

or, equivalently, (a more convenient form for us)

$$U_{N+n+1} = C_N^{N+n}U_N - \sum_{m=N}^{N+n} C_{m+1}^{N+n}[(A/2C)((W_{m+1}-W_N) - (W_m-W_N)) \tag{7.1}$$
$$+ AB(\theta)C^2((t_{m+1}-t_N) - (t_m-t_N))].$$

For the moment, let us consider <u>only</u> the <u>sum</u> in (7.1) involving the W_i. Denoting that sum by $\tilde{I}_N^{N+n}$ and rewriting it by collecting the coefficients of each W_i, yields

---

[+]It is precisely the difficulty of proving tightness of $\{U_n\}$ when (A5.6) is relaxed, that forces us to require (A5.6). See the last Section, where relaxations of the condition are discussed.

$$\tilde{I}_N^{N+n} = \sum_{m=N}^{N+n} [C_m^{N+n} - C_{m+1}^{N+n}] (A/2C) (W_m - W_N) + (A/2C) (W_{N+n+1} - W_N).$$

Using $C_m^{N+n} - C_{m+1}^{N+n} = -C_{m+1}^{N+n} (\bar{K}\Delta t_m + \varepsilon_{3,m}\Delta t_m)$, where $\bar{K} = \bar{K}_1$ or $\bar{K}_2$ according to the case of (A5.9), yields $\tilde{I}_N^{N+n} = I_N^{N+n} +$ $+ \hat{I}_N^{N+n} + (A/2C) (W_{N+n+1} - W_N)$, where

$$I_N^{N+n} = \sum_{m=N}^{N+n} C_N^{N+n} (C_N^m)^{-1} (-\bar{K}\Delta t_m) (A/2C) (W_m - W_N)$$

$$\hat{I}_N^{N+n} = \sum_{m=N}^{N+n} C_N^{N+n} (C_N^m)^{-1} (-\varepsilon_{3,m}\Delta t_m) (A/2C) (W_m - W_N).$$

Now, for each N, define $C_N(\cdot)$ on $[0,\infty)$ by

$$C_N(t) = C_N^{N+n} \quad \text{on} \quad [t_{N+n} - t_N, \; t_{N+n+1} - t_{N+n}).$$

Since $\sum_{n=N}^{\infty} (\Delta t_n)^2 \to 0$ as $N \to \infty$, we have that $C_N(t) \to \exp - \bar{K}t$, as $N \to \infty$, uniformly (w.p.1) on finite time intervals.

For each N, define the interpolations $I_N(\cdot), \hat{I}_N(\cdot), \tilde{I}_N(\cdot)$ of the sequences $\{I_N^{N+n}\}$, $\{\hat{I}_N^{N+n}\}$, $\{\tilde{I}_N^{N+n}\}$, by e.g.,

$$I_N(t) = I_N^{N+n} \quad \text{on} \quad [t_{N+n} - t_N, t_{N+n+1} - t_N).$$

It is not hard to show that $\{\hat{I}_N(\cdot)\}$ is tight on $D^r[0,\infty)$ and goes to the "zero" process weakly, as $N \to \infty$. Thus, we ignore it henceforth. Define

$$\tilde{J}_N(t) = -\int_0^t C_N(t^-)C_N^{-1}(s)\bar{K}(A/2C)W^N(s)ds + (A/2C)W^N(t)$$

$$J_N(t) = -\int_0^t (\exp -\bar{K}(t-s))\bar{K}(A/2C)W^N(s)ds + (A/2C)W^N(t).$$

It is also not hard to show that

$$\{\tilde{J}_N(t) - J_N(t)\}, \quad \{\tilde{J}_N(t) - I_N(t)\}$$

are each tight on $D^r[0,\infty)$ and tend to the "zero" processes weakly as $N \to \infty$. A similar development can be made for the sum in (7.1) which involves the $(t_m - t_N)$. Putting all the foregoing together and adding the neglected terms, we have

$$\begin{aligned} U^N(t) = {} & (\exp -\bar{K}t)U^N(0) \\ & + \int_0^t (\exp -\bar{K}(t-s))\bar{K}(A/2C)W^N(s)ds \\ & - (A/2C)W^N(t) \\ & + \int_0^t (\exp -\bar{K}(t-s))\bar{K}(AB(\theta)C^2)s\,ds - (AB(\theta)C^2)t \\ & + \bar{\varepsilon}_N(t), \end{aligned} \tag{7.2}$$

where $\bar{\varepsilon}_N(\cdot)$ is a process which tends to the zero process weakly, as $N \to \infty$. Note that if a subsequence of $\{U^N(0), W^N(\cdot)\}$ converges weakly, so does the subsequence of $\{U^N(\cdot)\}$, and the limit process has continuous paths w.p.1.

3°. Suppose that $W^N(\cdot)$ were tight on $D^r[0,\infty)$ and that any weak limit is a Wiener process with covariance $\Sigma(\theta)t$. Then the limit of $\{U^N(\cdot)\}$ corresponding to any weakly convergent subsequence of $\{W^N(\cdot),U^N(0)\}$ is equivalent (in law on $D^r[0,\infty)$) to the process given by (7.2) with $U^N(0),W^N(\cdot)$ replaced by some $\tilde{U}(0),\tilde{W}(\cdot)$, where $\tilde{W}(\cdot)$ is a Wiener process with covariance $\Sigma(\theta)$. The theorem follows from this, since the resulting right hand side of (7.2) solves (5.4) (as an integration of (7.2) by parts will show).

4°. Thus, we only need to prove the first sentence of 3°. We use Theorem 2 of Scott [14] with an appropriate change of notation. Scott's result is the following.

Let $\{v_n^N\}$ denote an array of scalar valued random variables where v_i^N, $i < n$, are $\mathscr{F}_n^N$ measurable, for some sequence of σ-algebras $\mathscr{F}_n^N$, which is non-decreasing in n, for each N. Let Σ be a positive real number. Let $E_{\mathscr{F}_i^N} v_i^N = 0$ w.p.1, and define $m_N(t) = \max\{i: E\sum_{j=0}^{i}(v_j^N)^2 \leq t\Sigma\}$, $t \in [0,\infty)$, all N. Define

$$V^N(t) = \sum_{i=0}^{m_N(t)} v_i^N. \tag{7.3}$$

Let (condition B of Scott's Theorem 2; $\xrightarrow{P}$ is convergence in probability)

$$(7.4)\qquad \sum_{i=0}^{m_N(t)} E\{(v_i^N)^2 \mid \mathcal{F}_i^N\} \xrightarrow{P} \Sigma t,$$

$$(7.5)\qquad \sum_{i=0}^{m_N(t)} E\{(v_i^N)^2 I_{\{|v_i^n| \geq \varepsilon\}} \mid \mathcal{F}_i^N\} \xrightarrow{P} 0, \text{ all } \varepsilon > 0,$$

as $N \to \infty$, for each $t < \infty$. Then $\{V^N(\cdot)\}$ is tight on $D[0,\infty)$, and the limit of any weakly convergent subsequence is a Wiener process with covariance Σt and mean zero. Actually Scott deals with $D[0,T]$ and $\Sigma = 1$, but his theorem is valid on $D[0,\infty)$ also (and for any $\Sigma > 0$, by a simple scaling).

Assume that $\Sigma(\theta) \neq 0$, for otherwise the result is trivially true. Let $\lambda \in R^r$ be such that $\lambda'\Sigma(\theta)\lambda > 0$. Identify $\mathcal{F}_i^N$ with $\mathcal{F}_{N+i}$, v_i^N with $\lambda'\delta W_i^N$, and Σ with $\lambda'\Sigma(\theta)\lambda$. Note that the convergence $X_n \to \theta$ and (A5.7) imply (7.4). Condition (A5.8) implies (7.5). Thus, by Scott's theorem $\{V^N(\cdot)\}$ is tight and converges weakly to a Wiener process with mean zero and covariance $(\lambda'\Sigma(\theta)\lambda)t$, as $N \to \infty$. Note that

$$E \sum_{j=0}^{i} (v_i^N)^2 = E \sum_{j=0}^{i} (\lambda' E\xi_{N+j}\xi_{N+j}'\lambda)\Delta t_{N+j},$$

and $E\xi_{N+j}\xi_{N+j}' \to \Sigma(\theta)$ as $N, j \to \infty$. Thus, the result remains true if $E|\lambda'\xi_{N+j}|^2 = E(v_j^N)^2$ is replaced by $\lambda'\Sigma(\theta)\lambda$ in the definition of $m_N(t)$. With this change, the definition of $V^N(\cdot)$ is the same as that of $\lambda'W^N(\cdot)$; thus $\lambda'W^N(\cdot)$ converges weakly to a Wiener process with mean zero and covariance $\lambda'\Sigma(\theta)\lambda$.

Since the result of the last paragraph is true for each λ

(if $\lambda'\Sigma(\theta)\lambda = 0$, then the $W^N(\cdot)$ converge to the "zero" process), we can conclude that $\{W^N(\cdot)\}$ is tight on $D^r[0,\infty)$, and that it converges weakly to the desired R^r valued Wiener process, with covariance $\Sigma(\theta)$, degenerate or not. Q.E.D.

8. Exploitation of the "Correlation Structure" of the Limit Process.

In order to illustrate a possible useful application of the correlation properties of (5.4), consider a simple scalar case with $B(\theta) = 0$, $2C = 1$, $\beta = \alpha/3$. Write $F = F(\theta)$, $\Sigma(\theta) = \sigma^2$; let $\alpha = 1$ and let $\bar{K} \equiv AF - \beta > 0$. Then (in steady state)

$$(8.1) \qquad E\bar{U}(t) = 0, \quad E\bar{U}(t)\bar{U}(t+s) = V \exp - \bar{K}s, \quad s > 0,$$
$$V = A^2\sigma^2/2(AF-\beta).$$

V is minimized by $A = 2\beta/F \equiv A_0$.

In practice, F is not usually known. In order to reduce the sensitivity of the iterate sequence $\{X_n\}$ to "large initial noises", and to partially rectify the slow convergence that occurs when A is too small, it is common for an $A > A_0$ to be used (where, of course, F must be guessed). For similar reasons, an $\alpha < 1$ is sometimes used, and what follows also holds for the $\alpha < 1$ case (in which case the correlation (8.1) uses $\beta = 0$ and $\bar{K} = AF$). As A increase, $\bar{K}$ increases, and the process $\{(n+1)^\beta(X_n-\theta)\} = \{U_n\}$ behaves more "wildly". We will try to exploit this.

Let $b \in [0,1]$, $0 < t < \infty$ and define $q_N(t) = \max\{i: \sum_{j=0}^{i} \Delta t_{j+N} \leq t\}$, $q_N(0) = 0$. Using the fact that $n^{1/3}(X_n-\theta) \xrightarrow{D} N(0,V)$, together with the correlation structure of (8.1) yields, for large N,

$$(8.2) \qquad E[b(X_N-\theta) + (1-b)(X_{N+q_N(t)} - \theta)]^2$$

$$\approx V\left[\frac{b^2}{N^{2/3}} + \frac{2b(1-b)\exp-\bar{K}t}{N^{1/3}(N+q_N(t))^{1/3}} + \frac{(1-b)^2}{(N+q_N(t))^{2/3}}\right].$$

((8.2) holds as a limit relation if we multiply both sides by $N^{2/3}$, and let $N \to \infty$.)

By definition of $q_N(\cdot)$,

$$\sum_{i=N}^{N+q_N(t)} \Delta t_i \approx t \approx \log[N+q_N(t)] - \log N,$$

and $e^{-t}(N+q_N(t))/N \to 1$, as $N \to \infty$. Then (8.2) is approximately (the ratios tend to unity as $N \to \infty$)

$$(8.3) \qquad \frac{V}{N^{2/3}}\left[b^2+2b(1-b)\exp-(\bar{K}+1/3)t + (1-b)^2\exp-2t/3\right].$$

At $b = 0$, the derivative of (8.3) with respect to b is

$$(8.4) \qquad \frac{2V}{N^{2/3}}(\exp-t/3)[\exp-\bar{K}t-\exp-t/3].$$

At $A = A_0$, $\bar{K} = \beta = 1/3$, and (8.4) = 0. So, under "ideal" conditions (at least from an "asymptotic" point of view), the linear combination inside (8.2) does not yield an improved estimate. But if $A > A_0$, then (8.4) is < 0, suggesting that we can improve the estimate of θ at iterate $N + q_N(t)$, by using some linear combination of past iterates. Such ideas have a natural appeal, and such smoothing is sometimes used in practice, irrespective of the value of A; but, we see that it can be harmful, unless $A > A_0$ (or $\alpha < 1$), and the weights are carefully selected.

An open, and interesting, question in the general vector case is whether A can be selected (still guaranteeing $X_n \to \theta$ w.p.1), but such that $\bar{K}$ has some complex eigenvalues - the $\{X_n\}$ sequence would then exhibit "oscillations" on the average, which perhaps, could be exploited (via suitable smoothing of the $\{X_n\}$) to obtain better estimates of θ.

Calculations made on simple sample problems indicate that complex eigenvalues (often corresponding to the eigenvalues with the smallest real parts) occur quite frequently for the Lagrangian algorithm of Sections 3 and 9. Perhaps smoothing can be usefully applied there.

It should be clear from the foregoing, that our interpolated process" analysis of $\{X_n\}$ and $\{U_n\}$ is rather natural, and that it can yield much more insight into the properties of the sequences, and smoothed functionals of the sequence, than can the more standard analysis of only the random variables $\{X_n, U_n\}$.

9. Asymptotic Rate for the Lagrangian Method of Section 3.

The development is close to that for Theorem 5.1, and only an outline will be given.

Assumptions. Assume (A5.1) to (A5.8) and

(A9.1) There is a $\bar{\lambda}$ such that $\lambda_n \to \bar{\lambda}$ w.p.1.

(A9.2) If $q_i(\theta) = 0$, then suppose that $\bar{\lambda}^i > 0$. (Of course, if $q_i(\theta) < 0$, then $\bar{\lambda}^i = 0$.)

(A9.3) Each $q_i(\cdot)$ has continuous and bounded first and second derivatives at θ.

(A9.4) $Q(\theta)$ (see Sec. 3) is of full rank.

Let a_n, c_n be as in Section 5, and let $b_n = \text{diag}(b^1,\ldots,b^s)/(n+1)^\alpha = \text{diag}(b_n^1,\ldots,b_n^s)$, where $b^i > 0$, $i \leq s$. Define

$$Q_i(\theta) = \{\partial^2 q_i(x)/\partial x^k \partial x^j,\ k,j = 1,\ldots,r\} \quad \text{at} \quad x = \theta, \quad i = 1,\ldots,s.$$

Define $\bar{K}_0 = A(F(\theta) + \sum_i \bar{\lambda}^i Q_i(\theta))$ and

$$\bar{K}_2 = \begin{bmatrix} \bar{K}_0 & AQ(\theta) \\ -BQ'(\theta) & 0 \end{bmatrix}.$$

(A9.5) Let $\alpha < 1$ ($\alpha = 1$, resp.) and let the eigenvalues of $\bar{K}_2$ ($\bar{K}_1 = \bar{K}_2 - \beta I$, resp.) have positive real parts.

(A9.6) Let θ satisfy the Kuhn-Tucker condition $f_x(\theta) + \sum_i \bar{\lambda}^i q_{i,x}(\theta) = 0$.

Remark. The $\lambda_n \to \bar{\lambda}$ convergence was not proved in [7], but it appears from our simulations that convergence of $\{\lambda_n\}$ occurs quite frequently. (A9.2) is often assumed in the deterministic case; it simply says that the "economic price" of an "active" resource is positive at the optimal point.

If $q_i(\theta) < 0$, then $q_i(X_n) \to q_i(\theta) < 0$ and (3.1), the convergence $X_n \to \theta$, and divergence of $\sum_n b_n$ together imply that $\lambda_n^i = 0$ for all large n. We will henceforth ignore q_i if $q_i(\theta) < 0$. We can and will assume that all s constraints are active at θ - with no loss of generality. The linear independence of the $q_{i,x}(\theta)$ is also commonly assumed in the analysis of deterministic algorithms.

If $\bar{K}_0$ is positive definite and A and B are diagonal with the same constant diagonal elements, then (using A9.4) Polyak [21, proof of Theorem 1] implies (A9.5).

Development of the Algorithm. Owing to the remarks above and since we are only concerned with large n, we can write (3.1) as $\lambda_{n+1}^i = \lambda_n^i + b_n^i q_i(X_n)$, $i = 1,\ldots,s$. Define $\delta\lambda_n = \lambda_n - \bar{\lambda}$. Then, using (A9.6), we can write

$$\begin{pmatrix}\delta X_{n+1}\\ \delta\lambda_{n+1}\end{pmatrix} = \begin{pmatrix} I_r - a_n(\bar{K}_0+\varepsilon_{1,n}), & -a_n(Q(\theta) + \varepsilon_{2,n}) \\ b_n(Q'(\theta) + \varepsilon_{3,n}) & I_s \end{pmatrix} \begin{pmatrix}\delta X_n\\ \delta\lambda_n\end{pmatrix}$$

(9.1)

$$- \begin{pmatrix} a_n(B(\theta)c_n^2+\varepsilon_{4,n}c_n^2) \\ 0 \end{pmatrix} - \frac{a_n}{2c_n}\begin{pmatrix}\xi_n\\ 0\end{pmatrix},$$

where I_t = identity in R^t.

The terms $\varepsilon_{i,n}$ all depend only on X_n and λ_n and tend to 0 w.p.1 as $n \to \infty$, by the convergence (of $\{X_n,\lambda_n\}$) and smoothness (on $f(\cdot),q(\cdot)$) assumptions.

Define $U_n = (n+1)^{\beta}\begin{pmatrix}\delta X_n\\ \delta\lambda_n\end{pmatrix}$. Following the method by which (5.2) was obtained from (5.1), we get (the $\bar{\varepsilon}_{i,n}$ have the properties of the $\varepsilon_{i,n}$ above)

$$U_{n+1} = [I + \frac{\beta}{(n+1)} I - (n+1)^{-\alpha}(\bar{K}_2+\bar{\varepsilon}_{1,n})]U_n$$

(9.2)
$$- (n+1)^{-\alpha}(n+1)^{\beta-2\gamma}\begin{pmatrix} AB(\theta)C^2 + \bar{\varepsilon}_{2,n} \\ 0 \end{pmatrix}$$

$$- (n+1)^{-\alpha/2}(n+1)^{-\alpha/2+\beta+\gamma}(A/2C)\begin{pmatrix}\xi_n\\ 0\end{pmatrix} (1+O(\tfrac{1}{n})).$$

Define $\Delta t_n = (n+1)^{-\alpha}$, t_n, $\delta W_n(\cdot)$ and $W^N(\cdot)$ as in Section 5. Let $U^N(\cdot)$ denote the function which equals U_{N+n} on $[t_{N+n}-t_N, t_{N+n+1}-t_N)$ (analogously to the definition of $W^N(\cdot)$. Set $\beta = 2\gamma = \alpha/3$. Then (9.2) can be rewritten as ($\bar{K} = \bar{K}_1$ or $\bar{K}_2$ according to whether $\alpha = 1$ or $\alpha < 1$)

$$U_{n+1} = [I - \Delta t_n(\bar{K}+\bar{\varepsilon}_{3,n})]U_n - \Delta t_n \begin{Bmatrix} AB(\theta)C^2 \\ 0 \end{Bmatrix} - (A/2C)\begin{pmatrix} \delta W_n \\ 0 \end{pmatrix} + \Delta t_n \bar{\varepsilon}_{4,n}. \tag{9.3}$$

Let $\bar{W}(\cdot)$ denote a standard R^r valued Wiener process, and let $\bar{U}(\cdot)$ be the (stationary process) solution to

$$d\bar{U}(t) = -\bar{K}\bar{U}(t)dt - \begin{pmatrix} AB(\theta)C^2 \\ 0 \end{pmatrix} dt - \begin{pmatrix} \frac{A}{2C}\Sigma^{1/2}(\theta)d\bar{W}(t) \\ 0 \end{pmatrix}. \tag{9.4}$$

Theorem 9.1. Assume (A9.1) to (A9.6), (A5.1) to (A5.8), and let $a_n = A/(n+1)^{\alpha}$, $b_n = B/(n+1)^{\alpha}$, $c_n = C/(n+1)^{\gamma}$, where $C > 0$, B is diagonal with positive elements, and A is positive definite. Let $\beta = 2\gamma = \alpha/3$. Then $\{U^N(\cdot), W^N(\cdot)\}$ is tight on $D^{2r+s}[0,\infty)$, and any weak limit of the $\{U^N(\cdot)\}$ has the law of the (stationary process solution) $\bar{U}(\cdot)$ in (9.4).

Remarks. The proof is almost the same as that of Theorem 5.1 and is omitted.

Owing to the presence of the "multiplier dynamics" in (9.4), it is more likely that $\bar{K}_i$ will have some complex eigenvalues. The consequent oscillations (of the correlation function) should be exploitable, via ideas such as those in Section 7, to yield a smoothed sequence of estimators which are better than $\{X_n\}$. Indeed, such a possibility justifies our point of view concerning the advantages of studying weak convergence of the processes $U^N(\cdot)$ to $\bar{U}(\cdot)$, over the simpler convergence in distribution of the R^r valued sequence $\{U_n\}$.

10. Asymptotic Rates for the Equality Constrained Algorithm of Section 4.

We will need the assumptions:

(A10.1) The $\phi_i(\cdot)$ have two continuous derivatives at θ.

(A10.2) The $\phi_{i,x}(\theta)$, $i = 1,\ldots,s$ (the rows of $\Phi(\theta)$) are linearly independent.

Let $v(y)$ (with components $v_1(y),\ldots,v_r(y)$) denote the vector $\pi(y)f_x(y)$. Let $(\pi f_x(y))_x$ denote the matrix

$$\begin{bmatrix} v'_{1,x}(y) \\ \vdots \\ v'_{r,x}(y) \end{bmatrix} = \begin{bmatrix} \partial v_1(y)/\partial x_1 & \cdots & \partial v_1(y)/\partial x_r \\ \vdots & & \vdots \\ \partial v_r(y)/\partial x_1 & \cdots & \partial v_r(y)/\partial x_r \end{bmatrix} \equiv \begin{bmatrix} w_1(y),\ldots,w_r(y) \end{bmatrix}$$

$$= \begin{bmatrix} w_{11}(y) & & w_{r1}(y) \\ \vdots & \cdots & \vdots \\ w_{1r}(y) & & w_{rr}(y) \end{bmatrix}.$$

Define

$$\bar{K}_0 = (\pi f_x(\theta))_x + k\Phi'(\theta)\Phi(\theta),$$

(A10.3I)[+] Let $\alpha = 1$, $\beta = 2\gamma = 1/3$. Let the eigenvalues of $\bar{K}_1 = A\bar{K}_0 - \beta I$ have positive real parts.

or

[+]See the discussion after Theorem 10.1, and, in particular, the representation (10.8) for $\bar{K}_0$.

(A10.3II) Let $\alpha < 1$, $\beta = 2\gamma = \alpha/3$. Let the eigenvalues of $\bar{K}_2 = A\bar{K}_0$ have positive real parts.

(A10.4) θ satisfies the necessary condition for a constrained minimum, $\pi(\theta)f_x(\theta) = 0$.

Let $\Sigma_\pi(X_n)$ denote the covariance (given $\mathcal{B}_n$) of the projection of ξ_n onto the orthogonal complement of the span of $\phi_{1,x}(X_n),\ldots,\phi_{s,x}(X_n)$. We use the terminology of Section 5, except that $\{X_n\}$ is given by (4.1). Under (A5.7), and the convergence (A5.4), there is a matrix $\Sigma_\pi(\theta)$ such that $\mathrm{cov}[\pi(X_n)\xi_n \mid \mathcal{B}_n] \to \Sigma_\pi(\theta)$ w.p.1, as $n \to \infty$.

Theorem 10.1. Assume (A5.1) to (A5.8) and (A10.1) to (A10.4). Define $U_n = (n+1)^\beta(X_n - \theta)$. Then $\{U^n(\cdot), W^N(\cdot)\}$ is tight on $D^{2r}[0,\infty)$, and there is a standard Wiener process $\bar{W}(\cdot)$ such that any weak limit of $\{U^N(\cdot)\}$ has the (stationary process solution) law of the $\bar{U}(\cdot)$ in (10.1), where $\bar{K} = \bar{K}_1$ or $\bar{K}_2$, according to the case of A10.3.

$$(10.1)\qquad d\bar{U}(t) = -\bar{K}\bar{U}(t)dt - A\pi(\theta)B(\theta)C^2dt - (A/2C)\Sigma_\pi^{1/2}(\theta)d\bar{W}(t).$$

Remark. The proof is very close to that of Theorem 5.1 and is omitted. We remark only on the expansion of (4.1), and on the conditions. The $\bar{\varepsilon}_{i,n}, \varepsilon_{i,n}$ have the same meaning as in Section 5.

With $X_n - \theta = \delta X_n$, we can write

$$\delta X_{n+1} = \delta X_n - a_n[\pi(X_n)\{f_x(X_n) + B(\theta)c_n^2 + \varepsilon_{1,n}c_n^2 + \varepsilon_{2,n}\delta X_n + \xi_n/2c_n\} + k\Phi'(\theta)\Phi(\theta)\delta X_n + \varepsilon_{3,n}\delta X_n]. \tag{10.2}$$

Using (A10.4) and the smoothness assumptions on $f(\cdot)$ and $\phi(\cdot)$, we get

$$\pi(X_n)f_x(X_n) = (\pi f_x(\theta))_x\delta X_n + \varepsilon_{4,n}\delta X_n. \tag{10.3}$$

Now, using (10.2), (10.3), the values of α,β,γ in A10.3((Case I or II), and a development of $(n+2)^{\beta}$ such as used in Theorem 5.1, we get (analogously to the unconstrained case)

$$U_{n+1} = [I - \Delta t_n(\bar{K}+\bar{\varepsilon}_{1,n})]U_n - \Delta t_n A\pi(\theta)B(\theta)C^2 - \sqrt{\Delta t_n}\,(A/2C)\pi(\theta)\xi_n - \Delta t_n(\bar{\varepsilon}_{2,n}+\xi_n O(\tfrac{1}{n})). \tag{10.4}$$

Using (10.4), the proof goes as it does for Theorem 5.1.

Remark on (A10.3). Without loss of generality, suppose that $\theta = 0$. Let $T(y)$ denote the tangent line or hyperplane to the curve or surface $\{x: \phi(x) = 0\}$ at y, and $T_0(y)$ the orthogonal complement to $T(y)$. In general $T_0(0) \supset \text{span}\{\phi_{i,x}(0),\ i \leq s\}$. We make the additional assumption that $T_0(0) = \text{span}\{\phi_{i,x}(0),\ i \leq s\}$. Let

$x = (x_1,\ldots,x_r)$ denote the generic point in R^r. With no loss of generality (and some gain in insight), we can assume that the basis is such that $x_1,\ldots,x_s$ and $x_{s+1},\ldots,x_r$ form bases for $T_0(0)$ and $T(0)$, resp. Using the last three sentences and (A10.2), we have that there is a non-singular $(s\times s)$ matrix $\tilde{\Phi}$ such that

$$\Phi'(0)\Phi(0) = \begin{bmatrix} \tilde{\Phi}'\tilde{\Phi} & 0 \\ 0 & 0 \end{bmatrix}. \tag{10.5}$$

Let N^ε denote an ε-neighborhood of $0 \in R^r$. There are differentiable functions $\ell_1(\cdot),\ldots,\ell_s(\cdot)$ on R^{r-s} such that if $x \in N^\varepsilon \cap \{x:\phi(x) = 0\} \equiv N^\varepsilon_\phi$, and ε is small enough, then

$$x_i = \ell_i(x_{s+1},\ldots,x_r), \quad i = 1,\ldots,s.$$

Henceforth, assume that ε is "sufficiently" small.

We will now develop a representation for $(gf_x(0))_x$. Let δ denote a small real number. By the definition of the tangent plane or line $T(0)$, $\ell_i(e_j\delta) = O(\delta^2)$, $j = s+1,\ldots,r$. Consequently,

for $j \geq s$, the smoothness of $\phi(\cdot), \pi(\cdot)$ and $f_x(\cdot)$ implies that

$$(10.6) \qquad \pi(e_j\delta + \sum_{i=1}^{s} e_i \ell_i(e_j\delta)) \cdot f_x(e_j\delta + \sum_{i=1}^{s} e_i \ell_i(e_j\delta))$$

$$= \pi(e_j\delta) f_x(e_j\delta) + O(\delta^2).$$

By the definition of $w_i(0)$,

$$(10.7) \qquad \lim_{\delta \to 0} \pi(e_i\delta) f_x(e_i\delta)/\delta = w_i(0).$$

Recall that $\pi(0) f_x(0) = 0$. Note that for each i, the vector

$$\pi(\delta e_i) f_x(\delta e_i) = [\text{projection of } f_x(\delta e_i) \text{ onto the tangent plane } T(\delta e_i)]$$

has components $O(\delta)$ on $T(0)$ and $O(\delta^2)$ on $T_0(0)$. Thus, by (10.7), $w_i(0)$ is in $T(0)$ and, hence, has the form $w_i(0) = \begin{bmatrix} 0 \\ g_i \end{bmatrix}$ for some $r - s$ vector g_i.

We will examine further the term $\overline{F}_\phi \equiv [g_{s+1}, \ldots, g_r]$. Define the function $\overline{f}(\cdot)$ on $N^\varepsilon \cap T(0)$ by

$$\overline{f}(x_{s+1}, \ldots, x_r) = f(\ell_1(x_{s+1}, \ldots), \ldots, \ell_s(x_{s+1}, \ldots), x_{s+1}, \ldots, x_r).$$

Note that the vector of the last r-s components of the left side of (10.6) is the gradient of $\bar{f}(\cdot)$ at $e_j\delta$ (modulo a term of order $O(\delta^2)$). This fact, together with (10.6) and (10.7) imply that $\bar{F}_\phi$ is the Hessian matrix of $\bar{f}(\cdot)$ at 0. Finally, $\bar{K}_0$ takes the form

$$\bar{K}_0 = \left[\begin{array}{c|c} k\tilde{\Phi}'\tilde{\Phi} & 0 \\ \hline g_1,\ldots,g_s & \bar{F}_\phi \end{array}\right]. \tag{10.8}$$

Thus (A10.3II) holds if the eigenvalues of the Hessian $\bar{F}_\phi$ have strictly positive real parts and $A = \text{diagonal}(a,a,\ldots)$, $a > 0$. In any case, the representation (10.8) clarifies the meaning of the condition (A10.3).

Remark of the value of k. If $\alpha = 1$, we require that the real parts of the eigenvalues of

$$A\bar{K}_0 - \beta I$$

be positive. This requires that $k \geq$ some minimum positive value.

Certain deterministic algorithms [15] also require a minimum value of k, for convergence. Here, convergence occurs for any k. But, if $\alpha = 1$, and k is too small, the rate (β) will not be $\alpha/3 = 1/3$, but something less, something which also seemed to hold in our experiments.

11. Extensions.

Assumption (A5.6) was used, because we were not otherwise able to prove tightness of $\{U_n\}$ in any generality. Theorem 11.1 follows from the proof of the previous theorems. It is much less difficult to find reasonable conditions under which $W^n(\cdot)$ converges weakly to a Wiener process. After the theorem statement, we will comment on this.

Theorem 11.1. Assume the conditions of Theorem 5.1 (or of (9.1, 10.1), except for (A5.6), and suppose that $\{U_n, W^n(\cdot)\}$ are tight, and $\{W^n(\cdot)\}$ converges weakly to a Wiener process with mean 0, and covariance $\Sigma_0(\theta)$. Then the conclusions of the theorems still hold.

Remarks. Our remarks will be confined to the unconstrained case, for the others are treated similarly. The $a_n\bar{\varepsilon}_n$ in (5.2) causes no problem, the difficulty in proving tightness of $\{U_n\}$ on R^r has been due to the $\varepsilon_{1,n}$ term, although it is "sure" to be eliminated eventually. Indeed, if tightness of $\{U_n\}$ can be shown, then even (A5.4) can be replaced by the weaker convergence in the theorem of Kushner [10].

Define $q_N(t) = 0$ on $[0,\Delta t_N)$, $q_N(t) = i$ on $[\Delta t_N+\ldots+\Delta t_{N+i-1}, \Delta t_N+\ldots+\Delta t_{N+i})$. Then

$$W^N(t) = \sum_{i=0}^{q_N(t)-1} \xi_{N+i}(\Delta t_{N+i})^{1/2}.$$

Tightness and convergence of $\{W^n(\cdot)\}$. Let us drop (A5.6) to (A5.8). We will replace them by the "reasonable" conditions (A11.1) to (A11.3). According to Billingsley [16], Theorem 15.5, if $\{W^N(\cdot)\}$ satisfies a criterion "similar" to that used to prove tightness[+] on $C^r[0,\infty)$, then it will be tight on $D^r[0,\infty)$, and all limits will be continuous w.p.1. In particular, by [16], Theorems 15.5, 12.3 and 12.2, and the definition of $W^N(\cdot)$, this holds if there is a real K such that

$$(11.1) \qquad E|W_{n+m} - W_n|^4 \le K\Big|\sum_{i=n}^{n+m-1} \Delta t_i\Big|^2, \quad \text{all} \quad n,m > n.$$

Equation (11.2) is equivalent to (11.1), where $t_n^N = t_{N+n} - t_N$.

$$(11.2) \qquad E|W^N(t^N_{n+m}) - W^N(t^N_n)|^4 \le K\Big|\sum_{i=n}^{n+m-1} \Delta t_i^N\Big|^2, \text{ all } N,n, m > n.$$

Just to simplify the notation in the following development, we assume that the ξ_n are scalar valued.

Let $\mathcal{F}_n$ be the σ-algebra determined by $X_0,\ldots,X_{n+1}$, $\xi_0,\ldots,\xi_n$. Assume

---

[+] $C[0,\infty)$ = space of continuous functions on $[0,\infty)$ with the metric of uniform convergence on finite intervals.

(All.1) There is an integer $k \geq 0$ such that for all N,i,

$$|E_{\mathcal{B}_N}\xi_{N+i}| \leq (1+|\bar{\xi}_N|)\alpha_i^N, \text{ where } \alpha_i^n \text{ are real quantities satisfying } \sum_{i=0}^{q_N(t)} \alpha_i^N(\Delta t_{N+i})^{1/2} \to 0 \text{ as } N \to \infty, \text{ for each } t, \text{ where we define } \bar{\xi}_N = \sum_{\ell=0}^{k} |\xi_{N-\ell}|.$$

Let $R(\theta;\ell)$ and $\beta_{i,\Delta}^N$ be real quantities such that

$$\sum_{\ell} |R(\theta;\ell)| < \infty, \quad \sum_{i,j=1}^{q_N(t)} \beta_{i,j}^N(\Delta t_{N+i}\Delta t_{N+j})^{1/2} \to 0$$

as $N \to \infty$, for each t.

(All.2) $|E_{\mathcal{B}_N}\xi_{N+i}\xi_{N+i+\ell} - R(\theta,\ell)| \leq (1+|\bar{\xi}_N|^2)\beta_{i,i+\ell}^N$, $\ell \geq 0$, for some integer $k \geq 0$, and all N,i,ℓ.

(All.3) There is a bounded function $R(\cdot,\cdot,\cdot,\cdot)$ and real number K such that $|E\xi_i\xi_j\xi_k\xi_\ell| \leq R(i,j,k,\ell)$ and (where t $t+s$ are restricted to jump times of $W^N(\cdot)$)

$$\sum_{i,j,k,\ell=q_N(t)}^{q_N(t+s)-1} R(N+i,N+j,N+k,N+\ell)(\Delta t_{N+i}\Delta t_{N+j}\Delta t_{N+k}\Delta t_{N+\ell})^{1/2} \leq Ks^2.$$

Discussion of the conditions. The conditions can all be weakened, but we do not know their "best" form. (All.1) replaces (A5.6). It describes the type of mixing or summability condition that holds if the $\{\xi_m\}$ were generated by the solution to an equation such as

$$\xi_{n+i+1} + a_0\xi_{n+i} + a_1\xi_{n+i-1} \cdots + a_i\xi_n = \psi_n, \tag{11.3}$$

where the $\{\psi_n\}$ are independent, Gaussian, zero mean, and identically distributed, and the roots of $[\lambda^{i+1} + a_0\lambda^i + \ldots + a_i = 0]$ are all strictly interior to the unit circle.

Similarly, (A11.2) holds for (11.3). If the noise $\{\xi_m\}$ were a stationary process - with $E\xi_i\xi_{i+\ell} = R(\ell)$, then (A11.2) would read

$$|E_{\mathcal{B}_N}\xi_{N+i}\xi_{N+i+\ell} - E\xi_{N+i}\xi_{N+i+\ell}| \leq (1+|\bar{\xi}_N|^2)\beta^N_{i,i+\ell}.$$

The condition (A11.2) is a type of "asymptotic" stationarity (as $X_n \to \theta$) combined with a mixing condition. Conditions (A11.1) - (A11.2) are used to show that the limit of $\{W^N(\cdot)\}$ is a Wiener process, given tightness, and (A11.3), which implies (11.2), also implies tightness. Also, it holds for (11.3).

(A11.3) actually is not too restrictive. For example, it commonly occurs that there are functions $\bar{R}(\cdot,\cdot)$ such that $R(i,j,k,\ell) \leq \bar{R}(i,j)\bar{R}(k,\ell) + \bar{R}(i,\ell)\bar{R}(k,j) + \bar{R}(i,k)\bar{R}(j,\ell)$, and where $\bar{R}(i,j) \leq M\alpha^{|i-j|}$ for some $\alpha \in (0,1)$, and real M. Then the sum in (A11.3) is bounded above by

$$3M^2\left(\sum_{i,j=q_N(t)}^{q_N(t+s)-1} \alpha^{|i-j|}(\Delta t_{N+i}\Delta t_{N+j})^{1/2}\right)^2 \tag{11.4}$$

and, in turn, the bound in (A11.3) can be verified from (11.4). Note that (A11.3) implies that:

(11.5) For each t, there is an $M_1(t) < \infty$ such that $E|W^N(t)|^4 \leq M_1(t)$, all N.

Under (A11.3), (11.2) holds and $\{W^N(\cdot)\}$ is tight on $D[0,\infty)$, and the paths of any limit process are continuous w.p.1. Assume that $\{U_N\}$ is tight on R_0^r. Then, $\{U^N(\cdot)\}$ is tight also.

We will prove that the limit of $\{W^N(\cdot)\}$ is a Wiener process with covariance $R_0(\theta) = R(\theta;0) + 2\sum_{i=1}^{\infty} R(\theta;i)$. First, some estimates are needed. Let $\mathcal{B}_N(t)$ be the smallest σ-algebra which measures $\{W^N(s), U^N(s), s \leq t\}$, and, for notational simplicity, fix s,t, and set $m_N = N + q_N(t)$.

By (A11.1), and the definition of $W^N(t+s) - W^N(t)$,

$$(11.6)\quad |E_{\mathcal{B}_N(t)} W^N(t+s) - W^N(t)| \leq \sum_{i=q_N(t)}^{q_N(t+s)-1} (\Delta t_{N+i})^{1/2} \alpha_i^N (1+|\bar{\xi}_{m_N}|).$$

We can write

$$(11.7)\quad E_{\mathcal{B}_N(t)} (W^N(t+s) - W^N(t))^2 = \sum_{i,j=q_N(t)}^{q_N(t+s)-1} (\Delta t_{N+i} \Delta t_{N+j})^{1/2} E_{\mathcal{B}_N(t)} \xi_{N+i} \xi_{N+j}$$

$$= \sum_{i=q_N(t)}^{q_N(t+s)-1} \Delta t_{N+i} E_{\mathcal{B}_N(t)} \xi_{N+i}^2$$

$$+ 2 \sum_{\ell \geq 1} \sum_{i=q_N(t)}^{q_N(t+s)-\ell-1} (\Delta t_{N+i} \Delta t_{N+i+\ell})^{1/2} E_{\mathcal{B}_N(t)} \xi_{N+i} \xi_{N+i+\ell}.$$

By (A11.2), we can rewrite (11.7) as

$$R(\theta;0)\sum_{i=q_N(t)}^{q_N(t+s)-1}\Delta t_{N+i} + 2\sum_{\ell\geq 1}R(\theta;\ell)\sum_{i=q_N(t)}^{q_N(t+s)-\ell-1}(\Delta t_{N+i}\Delta t_{N+i+\ell})^{1/2}$$

(11.8)

$$+ F_N(t,s),$$

where the coefficients of the $R(\theta;0)$ or $2R(\theta;\ell)$ tend to s, as $N \to \infty$, and where

$$(11.9)\qquad |F_N(t,s)| < 2(1+|\bar{\xi}_{m_N}|^2)\sum_{i,j=q_N(t)}^{q_N(t+s)-1}\beta^N_{i,j}(\Delta t_{N+i}\Delta t_{N+j})^{1/2}.$$

Let $h(\cdot)$ be a real valued bounded continuous function on some Euclidean space R^{2q}, with $t_1,\ldots,t_q$ arbitrary real numbers $\leq t$. Let $\{U^N(\cdot),W^N(\cdot)\}$ denote a weakly convergent subsequence. Let the limit process be denoted by $\bar{U}(\cdot),\bar{W}(\cdot)$. By (11.6) and the uniform integrability (11.5) and (All.1),

$$(11.10)\qquad Eh(W^N(t_i),\ U^N(t_i),\ i \leq q)\ (W^N(t+s) - W^N(t)) \to 0,$$

By weak convergence, and the uniform integrability (11.4), the left side of (11.10) also tends to

$$Eh(\bar{W}(t_i),\bar{U}(t_i),\ i \leq q)(\bar{W}(t+s) - \bar{W}(t)),$$

which must thus equal 0. Similarly, (11.8), (11.9), (A11.2) and the weak convergence and uniform integrability (11.4) yield that

$$Eh(\bar{W}(t_i),\bar{U}(t_i),\ i \leq q)[(\bar{W}(t+s) - \bar{W}(t))^2 - R_0(\theta)s] = 0.$$

The last paragraph, together with the arbitrariness of $h(\cdot)$, t,s, $t_i \leq t$, and the continuity of $\bar{W}(\cdot)$, w.p.1, imply that $\bar{W}(\cdot)$ is a continuous martingale with quadratic variation $R_0(\theta)s$, hence it is the asserted Wiener process.

REFERENCES

[1] M. T. Wasan, Stochastic Approximation, Cambridge University Press, Cambridge, 1969.

[2] L. Ljung, "Convergence of Recursive Stochastic Algorithms", Report 7403, 1974, Lund Institute of Technology, Division of Automatic Control, Lund, Sweden.

[3] L. Ljung, T. Soderstrom, I. Gustavsson, "Counterexamples to General Convergence of a Commonly Used Recursive Identification Method", IEEE Trans. on Automatic Control, AC-20, 1975, pp. 643-652.

[4] V. Fabian, "Stochastic Approximation of Constrained Minima", Proc. 4th Prague Conference on Statistical Decision Theory and Information Theory", 1966, pp. 277-289.

[5] H. J. Kushner, H. T. Gavin, "Stochastic Approximation-like Algorithms for Constrained Systems: Algorithms and Numerical Results", IEEE Trans. on Aut. Control, AC-19, 1971, pp. 349-357.

[6] H. J. Kushner, "Stochastic Approximation Algorithms for Constrained Optimization Problems", Ann. Statist., 2(1974), pp. 713-723.

[7] H. J. Kushner, E. Sanvicente, "Stochastic Approximation for Constrained Systems with Observation Noise on the System and Constraints", Automatica, 11(1975), pp. 375-380.

[8] H. J. Kushner, E. Sanvicente, "Penalty Function Methods for Constrained Stochastic Approximation", J. Math. Anal. and Applic., 46(1974), pp. 499-512.

[9] H. J. Kushner, M. L. Kelmanson, "Stochastic Approximation Algorithms of the Multiplier Type for the Sequential Monte Carlo Optimization of Constrained Systems", SIAM J. on Control, 14, August 1976.

[10] H. J. Kushner, "General Convergence Results for Stochastic Approximations via Weak Convergence Theory", to appear J. Math. Anal. and Applic.

[11] A. Miele, E. G. Cragg, R. R. Iyer, A. V. Levy, "Use of Augmented Penalty Function in Mathematical Programming Problems", Part I; J. Optimiz. Theory and Applications, 8(1971), pp. 115-130.

[12] V. Fabian, "On Asymptotic Normality in Stochastic Approximation", Ann. Math. Statist., 39(1968), pp. 1327-1332.

[13] J. Sacks, "Asymptotic Distribution of Stochastic Approximation Procedures", Ann. Math. Statist., 29(1958), pp. 273-405.

[14] D. J. Scott, "Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorokhod Representation Approach", Adv. in Appl. Prob. 5(1973), pp. 119-137.

[15] D. I. Bertsekas, "Multiplier Methods: A Survey", Automatica, 12(1976), pp. 133-145.

[16] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[17] W. Whitt, "A Guide to the Applications of Limit Theorems for Sequences of Stochastic Processes", Oper. Res., 18(1970), pp. 1207-1213.

[18] H. J. Kushner, Probability Methods for Approximations in Stochastic Control and for Elliptic Equations, Academic Press, New York, to appear, late 1976 or early 1977.

[19] D. E. Iglehart, "Diffusion Approximations in Applied Probability", Math. of the Decision Sciences, Part II, Lectures in Appl. Math., 12(1968), Amer. Math. Soc., Providence, R.I..

[20] T. Lindvall, "Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0,\infty)$; J. Appl. Prob., 10(1973), pp. 109-121.

[21] B. T. Polyak, "Iterative Methods Using Lagrange Multipliers for Solving Extremal Problems with Constraints of the Equation Type", Zh. Vychisl. Mat. mat. Fiz., 10(1970), pp. 1098-1106 (Transl. pp. 42-52).

[22] H. J. Kushner, S. Lakshmivarahan, "Numerical studies for constrained stochastic approximation", to be submitted to IEEE Trans. on Automatic Control.

III. NUMERICAL STUDIES OF STOCHASTIC APPROXIMATION PROCEDURES FOR CONSTRAINED PROBLEMS

by

Harold J. Kushner[+], S. Lakshmivarahan[++]

August 1976

[+]Brown University, Providence, Rhode Island, Lefschetz Center for Dynamical Systems. Supported by the Air Force Office of Scientific Research AF-AFOSR 76-3063, National Science Foundation Eng 73-03846-A01, and Office of Naval Research NONR N000 14-76-C-0279.

[++]Brown University, Providence, Rhode Island. Lefschetz Center for Dynamical Systems. On leave from I.I.T., Madras, India. Supported by the Air Force Office of Scientific Research AF-AFOSR 76-3063 and the Office of Naval Research NONR N000 14-76-C-0279.

NUMERICAL STUDIES OF STOCHASTIC APPROXIMATION PROCEDURES FOR CONSTRAINED PROBLEMS

ABSTRACT

Four algorithms for stochastic approximation under equality and inequality constraints are discussed and described, together with numerical data and compariosns from numerous simulations. The algorithms work well, and exhibit some rather interesting behavior. Those based on "augmented Lagrangian" techniques are preferable to the "Lagrangian" methods, as in the deterministic case; the former methods seem to be quite robust and reliable. The study is the first (to the authors' knowledge) numerical study of such algorithms.

NUMERICAL STUDIES OF STOCHASTIC APPROXIMATION PROCEDURES FOR CONSTRAINED PROBLEMS

Harold J. Kushner and S. Lakshmivarahan

1. Introduction

The purpose of this paper is to discuss several recently developed algorithms for constrained stochastic approximation (SA), and to give some typical numerical results, taken from a large number of simulations. Most work on SA has dealt with the unconstrained case (see references in Wasan [1], and Ljung [2], for example). Nevertheless, there are numerous applications for the constrained problem - which has received relatively little attention; see Fabian [3], Kushner [4], Kushner and Gavin [5], Kushner and Sanvicente [6], [7], Kushner and Kelmanson [8], Kushner [9], [10]. Some of the many possible applications are described in [6].

References [3], [7] dealt with penalty type methods, references [4], [5] with stochastic analogs of the methods of feasible directions, [6] with a Lagrangian method, [8] with several methods of the "augmented" Lagrangian type and [9], [10] dealt with more general convergence proofs, and results on rates of convergence and exploited ideas in the theory of weak convergence of a sequence of probability measures - to obtain general results rather efficiently. While our understanding of the theoretical properties of the SA algorithms for

the constrained problem is progressing, virtually nothing is known about the numerical properties. Obviously, an understanding of the numerical problem is important for potential applications. But, it also points toward needed further theoretical development. As far as the authors are aware, this is the first paper (besides [5]) to discuss numerical results for constrained SA.

Reference [5] discussed some numerical results for the stochastic methods of feasible directions. While the results there were interesting, and exhibited some of the salient features of the stochastic problem, some problems were encountered with the behavior of the iterates on the boundary (the procedure seemed too inefficient there). These problems suggested the importance of investigating "dual" type methods. This paper is devoted to a description (with numerical data) of several such methods: there is little doubt that (if non-feasible iterates are allowable) their performance is considerably superior to the performance of the methods in [5].

In many respects, the behavior of the constrained algorithms is better than the behavior of the standard (unconstrained) SA. The iterates are bounded, and the constraints often force the iterates to remain near a boundary, thus further restricting the state space. It would appear, from our simulations, that there should be no more hesitation in using the "constrained" algorithms, than there would be for "unconstrained" SA. Also, there are, undoubtably, applications to areas such as identification, which have not yet been explored.

The convergence proofs for all the algorithms given here - appear elsewhere [6], [8]. The main concern of this paper is with the numerical properties, and with a description of the observed path behavior and its dependence on the parameters of the algorithm. Section 2 treats an equality constrained problem (algorithm 1 of [8]). Section 3 treats a Lagrangian method (from [6]), and Sections 4 and 5 treat the "augmented" Lagrangian algorithms 3 and 4, resp. of [8]. The algorithms worked surprisingly well.

All the reported simulations are on two dimensional problems, except for some results in Section 4, which concern a 5 dimensional problem.

2. An Equality Constrained Algorithm

The problem and the algorithm. The problem is to sequentially minimize $f(\cdot)$, under constraints $\phi_i(x) = 0$, $i = 1,\ldots,s$. The function $f(\cdot)$ will always be real valued, be defined on the Euclidean parameter space R^r, and have bounded second derivatives. The functions $\phi_i(\cdot)$ are real valued, with continuous first derivatives, and are assumed known. As is usual in SA, the form of $f(\cdot)$ is not assumed known. However, at any selected parameter setting (say x), a noise corrupted estimate of the performance, namely $f(x)$ + noise, can be observed. This provides our only information on $f(\cdot)$. In the simulations, the noise sequence will be independent, Gaussian and have mean zero.

Let $\{a_n, c_n\}$ denote real positive null sequences (the second being the n^{th} finite difference interval), k a positive

scalar, $P(x) = \sum_i |\phi_i(x)|^2$, e_i the unit vector in the i^{th} coordinate direction, and $\{X_n\}$ the sequence of estimates of the constrained minimum, which we always denote by θ. Define the matrix $\Phi(x) = [\phi_{1,x}(x),\ldots,\phi_{s,x}(x)]'$, where $'$ is transpose and the subscript x on $\phi_{i,x}(\cdot)$ denotes gradient. Let the operator $\pi(x)$ be defined by: $(I-\pi(x))y$ is the projection of a vector y onto $\text{span}[\phi_{1,x}(x),\ldots,\phi_{s,x}(x)]$; i.e., $\pi(x)y$ is the projection of y onto the tangent line or plane to the curve or surface $\{z: \phi(z) = \phi(x)\}$, at $z = x$. The vectors $\delta\overline{f}(x,c)$ and $\delta f(x,c)$ defined by

$$\delta f(x,c) = [\delta f^1(x,c),\ldots,\delta f^r(x,c)]' =$$

$$= \frac{1}{2c_n}\,[f(x+e_1c) - f(x-e_1c),\ldots]',$$

$$\delta\overline{f}(x,c) = \frac{1}{c_n}\,[f(x+e_1c) - f(x),\ldots]',$$

are the central difference and one-sided difference estimates of the gradient $f_x(x)$. Let the "noisy" gradient estimate (central difference) $\delta Y_n = (\delta Y_n^1,\ldots,\delta Y_n^r)'$ be defined by

$$\delta Y_n^i = \frac{[f(X_n+e_ic_n) - f(X_n-e_ic_n)]}{2c_n} + \frac{[\xi_{n,i}^+ - \xi_{n,i}^-]}{2c_n}$$

$$= \delta f^i(X_n,c_n) + \xi_n^i/2c_n\ ,$$

where $\xi_{n,i}^+, \xi_{n,i}^-$ are the observation noises in the measurements of performance at $X_n \pm e_ic_n$, resp.. Define $\delta\overline{Y}_n$ similarly, by $\delta\overline{Y}_n^i = \delta f^i(X_n,c_n) + \xi_n^i/c_n$, where now $\xi_n^i = \xi_{n,i}^+ - \xi_{n,i}^o, \xi_{n,i}^o$

being the observation noise in the performance measurement at X_n.

Algorithm 1 of [8] is a stochastic version of an augmented Lagragian algorithm of Miele, et. al [11], and takes the form (using central differences)[+]

$$X_{n+1} = X_n - a_n[\pi(X_n)\delta Y_n + \frac{k}{2} P_x(X_n)] \tag{2.1}$$

$$= X_n - a_n[\pi(X_n)\delta f(X_n,c_n) + \pi(X_n)\xi_n/2c_n + \frac{k}{2} P_x(X_n)].$$

In all the simulations of the paper, if the step size $|X_{n+1} - X_n|$ obtained from (2.1) was $> \frac{1}{2}$, we truncated it and defined X_{n+1} by ($|\cdot|$ is the Euclidean norm)

$$X_{n+1} = X_n + \frac{\text{step from (2.1)}}{2 \cdot |\text{step from (2.1)}|} .$$

Such a truncation is clearly necessary in any SA procedure, to assure that the sequence will not be oversensitive to large errors (noise, bias) or overreliance (resulting in steps of excessive size) on values of $f_x(x)$ or $\delta f(x,c)$ in the early stages of the procedure.

Under some additional conditions on a_n, c_n, ξ_n (which will always be satisfied here) $\{X_n\}$ converges w.p.1 to a point θ, which satisfies the necessary condition of the calculus for a constrained minimum (original proof in [8], weakened conditions in [9]).

[+]The algorithm in [8] uses $f_x(X_n)$ + noise for the observation, but (as for the other algorithms of the sequel), the finite difference form and proof require only small changes.

In the simulations to be reported on we used $f(\cdot)$ of the quadratic form

$$f(x) = (x_1+2.5)^2 + (x_2 - \tfrac{1}{2})^2/4 \tag{2.2}$$

and (only one constraint in 2-space variables)

$$\phi(x) = -\sin x_1 + x_2. \tag{2.3}$$

Actually, other constraints were studied, but (2.3) allows a fairly stringent test, and exhibition of the features of the algorithm - at least in R^2.

The finite difference estimate $\delta f(x,c)$ of $f_x(x)$ (for (2.3)) is unbiased. In order to add bias, without getting involved in a choice of a more complicated function, we used $\delta\bar{f}(x,c), \delta\bar{Y}_n$ in (2.1), in lieu of $\delta f(x,c), \delta Y_n$ - in all the simulations. Note, however, that the biases for a general, smooth, $f(\cdot)$ satisfy

$$\begin{aligned} |f_{x_i}(x) - \delta f^i(x,c)| &\approx |f_{x_i x_i x_i}(x)c^2|/3! \\ |f_{x_i}(x) - \delta \bar{f}^i(x,c)| &\approx |f_{x_i x_i}(x)c|/2!, \end{aligned} \tag{2.4}$$

and so our "trick" has created a somewhat larger bias (for small c_n) than we would get for a general smooth $f(\cdot)$. Thus, with use of a central difference estimate, the algorithm would actually perform better than indicated by the simulations.

The form of $\{a_n, c_n\}$. Let n_0 denote an integer (= 100 in all the simulations), let $\alpha, \gamma, \delta, A$ and C be positive real numbers with $\delta > \alpha$. Define the sequence $\{m_n\}$ as follows: $m_0 = m_1 = 1$; for $n \geq 2$, $m_n = m_{n-1} + 2$ if the angle formed by the 3 points (X_{n-2}, X_{n-1}, X_n) is less than 90°. Otherwise $m_n = m_{n-1}$. Set $a_n = A/m_n^{\delta}$, $n \leq n_0$; $a_n = A/m_n^{\alpha}$, $n > n_0$, $c_n = C/(n+1)^{\gamma}$.

The method of calculating a_n, c_n, m_n provides a kind of adaptation. If the angles are generally $\geq$ 90°, we "presume" that the sequence is moving more or less monotonically toward its goal. If the angle is < 90°, we presume that it was caused by either a noise effect or an overshoot, and increase m_n, hence decrease a_n. The increase in m_n is 2, since, in the pure noise situation, the angle will be less than 90° one-half of the time, and we want the average long term value of m_n to be n (which, indeed, it was - modulo reasonable statistical fluctuations).

The break in the rate of decrease of a_n, at $n = 100$, is for the following reason. Clearly, some form of adaptation is ncessary. The problem of choice of a_n, c_n in the constrained problem is much more difficult than it is in the unconstrained problem. However, experience with SA simulations indicates that it can be harmful if a_n decreases too rapidly in the initial phases. The value of a_n may become too small, while the iterates are still quite far from the region of the constrained minimum: a_n should not be allowed to decrease too

fast until it is clear that either overshoots are becoming a problem (which is not usually too serious when there is a step size limitation) or that noise effects begin to play a serious and steady role. The choice taken above, involving the n_0, and $\delta < \alpha$, was a simple (and effective within the context of the simulations) attempt to deal with this. The topic will be returned to in connection with the discussion below concerning the rate of convergence.

Theoretical rate of convergence.[+] Let $\pi(x)f_x(x) = (v_1(x),\ldots,v_r(x))'$, and $w(x) = [v_{1,x}(x),\ldots,v_{r,x}(x)]'$. The matrix $w(\theta)$ is (in the proper coordinate system) the Hessian matrix (at θ) of the function $f(\cdot)$, defined on the constrained surface. Define $\bar{K} = w(\theta) + k\Phi'(\theta)\Phi(\theta)$, let $\gamma < \alpha \leq 1$, $\beta = \min[2\gamma,\alpha/3]$, $\bar{\beta} = \min[\gamma,\alpha/4]$ and $B = (B_1,\ldots,B_r)$, $\bar{B} = (\bar{B}_1,\ldots,\bar{B}_r)$, where $B_i = (f_{x_ix_ix_i}(\theta)/3!)$ and $\bar{B}_i = f_{x_ix_i}(\theta)/2$. Assume, for simplicity (and it is also true in our simulations), that $\mathrm{Cov}[\xi_n \mid \xi_0,\xi_1,\ldots,\xi_{n-1}] \to \sigma^2 I$, for some constant σ^2; as $n \to \infty$. Define $\bar{U}^1,\bar{U}^2$ to be random vectors whose distributions are the stationary distributions of the solutions to (2.5a,b) resp., (which will exist, under the conditions below). Let (in this subsection only) $a_n = A/(n+1)^{\alpha}$, $c_n = C/(n+1)^{\gamma}$.

[+]This subsection quotes a result from [10], which is not yet published. The result helps to explain some of the numerical data and the problem of parameter selection, but the result is not critical here.

(2.5a) $dU^1(t) = -(A\bar{K}-\beta I)U^1(t)dt - AC^2Bdt - (A/2C)\sigma dw,$

(2.5b) $dU^2(t) = -A\bar{K}\ U^2(t)dt - AC^2Bdt - (A/2C)\sigma dw,$

where $w(\cdot)$ is a standard Wiener process.

The following result is in [10], and, after some additional work, can also be obtained in part from [12]. We do not list all the conditions for validity, but they all hold for our case. Here, A is either a scalar or a matrix.

I. Use the central difference in (2.1).

(Ia) Let $\alpha = 1$, and $\beta = 2\gamma = \alpha/3$, and assume that all the eigenvalues of $-(A\bar{K}-\beta I) \equiv -\bar{K}_1$ have negative real parts. Then $n^{\beta}(X_n-\theta) \xrightarrow{D} \bar{U}^1$ (in distribution).

(Ib) Let $\alpha < 1$, $\beta = 2\gamma = \alpha/3$, and let the eigenvalues of $-A\bar{K} = -\bar{K}_2$ have negative real parts. Then $n^{\beta}(X_n-\dot{\theta}) \xrightarrow{D} \bar{U}^2$.

(Ic) If $2\gamma < \alpha/3$, the noise is less important than the bias, and the above (Ia or b) holds with $\sigma = 0$ in (2.5). If $2\gamma > \alpha/3$, the bias is less important than the noise, and (Ia,b) holds with $B = 0$ in (2.5).

(Id) If (Ia) holds except for the eigenvalue criterion, but the eigenvalues of $-(A\bar{K}-bI)$ have negative real parts for some $b \in (0,\beta)$, then $n^{b}(X_n-\theta) \xrightarrow{D} 0$. There is an analogous version of (Ic).

II. Use the one sided difference $\delta\bar{Y}_n$ in (2.1). Then (Ia-d) continue to hold, with $A/C,\bar{\beta},\bar{B}$ and $\gamma > \alpha/4$ (or =) replacing $A/2C,\beta,B$ and $2\gamma > \alpha/3$ (or =), resp.

On the choice of A,C,α,γ,δ, returned. In Case I, $\alpha = 1$, $\gamma = \frac{1}{6}$ yields the best asymptotic rate, and $\alpha = 1$, $\gamma = \frac{1}{4}$ does so in Case II. The rate result can also be used as a basis for choosing A and C. But, note that selecting the parameters via a minimization of the normed asymptotic variance must be done with great care. It is not clear what "asymptotic" means, when the number of iterations is only in the hundreds, or is even less. Often using $\alpha = 1$ led to relatively poor results in the first few hundred iterations. The a_n decreased too fast, and often left the iterate sequence "stuck" away from the vicinity of θ. So, we prefer to use $\alpha < 1$. The value $\alpha = \frac{5}{6}$ seemed to yield reasonably good results. In the runs, with $\alpha = 1$, when there was a reasonably rapid (within the first 100 iterations) "initial" convergence to a small enighborhood of θ, then $\alpha = 1$ performed (as expected) slightly better than similar runs with $\alpha = \frac{5}{6}$; there was a little more "noise stability". When α was reduced below $\frac{5}{6}$, the behavior worsened (the "asymptotic" noise effects increased - without a compensating improvement in the initial behavior). We also settled on $\gamma = \frac{1}{6}$, which is between the requirements of I and II. Note that the simulations used $a_n = A/m_n^{\delta}$ or A/m_n^{α}.

It is difficult (and perhaps dangerous) to generalize from a few simulations. However, on the average (of 20 run sequences of 500 iterations each) for the case of Figure 1.1, where $k = 1$ and $A = \frac{1}{4}$, the value $\alpha = \frac{5}{6}$ did several percent better than did the value $\alpha = 1$; $\alpha = \frac{5}{6}$ always

did better when the convergence in the first 100 iterates was relatively poor. On the other hand, a series of 20 simulations with $\alpha = 1$, $A = \frac{1}{8}$ did (a few percent) better than did a series with $\alpha = \frac{5}{6}$, $A = \frac{1}{4}$ (the averages $\bar{f}(X_{500})$ were within .04 and .06 of the minimum value, and the x_{500} points were fairly close to the curve in both cases). It was found that $\alpha = \frac{5}{6}$ was a reasonable overall choice.

As is clear from the rate formula, the optimal (asymptotic) value of A depends on the Hessian of $f(\cdot)$ on the constrained surface and on $\Phi(\cdot)$ at θ. We took a relatively simple method for selecting A - by assuming that the problem was unconstrained, and making a rough estimate (from data) of the appropriate A for the unconstrained problem, and then using a slightly smaller value. In fact, the optimal (asymptotic) value of A for a one-dimensional unconstrained iteration in the x_1 direction is $2\beta/f''(\theta)$, provided that $\alpha = 1$, $\gamma = \frac{1}{6}$, $B = 0$. If $\alpha < 1$, then smaller values of A are preferable. Larger values did not perform as well.

The value $\delta = \frac{2}{3}$ was used, again chosen via trial and error. Larger values have the same liabilities as do larger values of α, and with smaller values, the cummulative noise effect was generally too great.

Note on the choice of k. In various deterministic forms of the agumented Lagrangian method, the algorithm yields $X_n \to \theta$ only if $k \geq$ minimum value > 0 [13]. Here $k > 0$ is arbitrary [8]. But, we see from the rate result that the rate depends on k, and may be low if k is too small. Such a phenomenon was observed in the simulations, and provided one of the

motivations for the derivation of the rate expressions. Of course, if we increase A, then k is effectively increased, as far as the rate is concerned.

Numerical data. See Figures 2.1-2.3, for some typical runs, each of 500 iterations. Note that $\sigma^2 = 2$, a fairly large value. We have $f(\theta) = .255$ and $\theta = (-2.7,-.43)$. When we observed $f_x(X_n)$ + noise in lieu of $\delta\overline{Y}_n = \delta\overline{f}(X_n,c_n)$ + noise/$2c_n$, the runs were excellent, converging rapidly and directly to θ, even with $\sigma^2 = 2$.

The figures illustrate the effect of k. Figures 2.1 and 2.2 are similar, except for the pronounced (expected) compression of the trajectory about the curve in Figure 2.2. The oscillations (and their orientation with respect to the constraint curve) in Figure 2.1 are caused by the fact that from iteration 50 on, the term $\pi(X_n)f_x(X_n) \approx \pi(X_n)\delta\overline{f}(X_n)$ is fairly small, and $\pi(X_n)\xi_n/c_n$ fairly large. Both vectors lie in the tangent line to the curve $\{y: \phi(y) = \phi(x)\}$ at $x = X_n$, and so the movement is largely random - and oscillates along the "tangent lines". The $kP_x(X_n)$ term pulls the iterate to the curve. (In Figures 2.1 to 2.3, we see something like a one-dimensional unconstrained SA - stretched into a helix to emphasize the rather random "oscillating" approach to θ.) If k is too small, then the movement to the curve is too slow, and since $a_n > 0$, the procedure virturally grinds to a halt (as happened

with $k = .1$). A similar phenomenon is illustrated by Figure 2.3, which we show to illustrate the phenomenon only - the run itself was worse than the average. It seems preferable, as in the deterministic case, to select the largest k for which the numerical process is reasonably well conditioned. Of course, k can (should) be adjusted in the course of the iterations.

Table 2.1 illustrates the effect of k. The numerical averages of $\phi(X_{500}), f(X_{500})$ and numerical variance of $f(X_{500})$ over 20 independent runs are listed. The data (besides k) is that of Figures 2.1 or 2.2.

	$\bar{\phi}(X_{500})$	$\bar{f}(X_{500})$	var $f(X_{500})$
$k = \frac{1}{2}$	-.0168	.3090	.00194
1	-.0126	.3046	.00228
4	-.0013	.3003	.00241

Table 2.1. The effect of k

The larger k are preferable, but not by much. As k decreased below $\frac{1}{2}$, the performance deteriorated seriously.

Let us return to the rate result. It can be shown [10] that one eigenvalue of $\bar{K}_2$ is proportional to k; of course - both are proportional to A. Let A be a scalar = $\frac{1}{4}$, and use the data of Figure 2.1. If $k = 2$, the eigenvalue proportional to k has value .909, and the other .305. Thus, k can be reduced to $2(.305)/(.909)$ before the convergence rate is dominated by the eigenvalue which is proportional to k. This is not inconsistent with our numerical observations.

3. A Lagrangian Algorithm

The algorithm. All undefined terms are as in the previous sections. Consider the Lagrangian algorithm of [6] for minimizing $f(\cdot)$ subject to inequality constraints $q_i(x) \leq 0$, $i = 1,\ldots,s$. Unless otherwise mentioned, the $q_i(\cdot)$ are known and continuously differentiable (also in Sections 4,5). Data on $f(\cdot)$ is, again, obtained only via noise corrupted observations. (The algorithm is valid if the $q_i(\cdot)$ are unknown, provided that noise perturbed observations of the constraint functions are taken; this will be commented on below.)

Let M_λ^i and M_x denote numbers such that $|\theta^i| \leq M_x$ and the multiplier for $q_i(\cdot)$ at θ is $\leq M_\lambda^i$. Let $\{b_n\}$ be a null sequence of positive real numbers, and $q(x) = (q_1(x),\ldots,q_s(x))'$. Define $\{X_n,\lambda_n\}$ by $\lambda_n = (\lambda_n^1,\ldots,\lambda_n^s)'$ and

$$\text{(3.1a)} \qquad \tilde{X}_n = X_n - a_n[\delta Y_n + \sum_i \lambda_n^i q_{i,x}(X_n)]$$

$$\lambda_{n+1}^i = \max[0,\min\{M_\lambda^i, \lambda_n^i + b_n q_i(X_n)\}],$$

$$\text{(3.1b)} \qquad X_{n+1}^i = \tilde{X}_n^i \quad \text{if} \quad |\tilde{X}_n^i| \leq M_x$$

$$= M_x \quad \text{if} \quad \tilde{X}_n^i > M_x$$

$$= -M_x \quad \text{if} \quad \tilde{X}_n^i < -M_x.$$

Again, the maximum size of $|X_{n+1} - X_n|, |\lambda_{n+1} - \lambda_n|$ was limited to $\frac{1}{2}$. In our simulations, we used $\delta\bar{f}(X_n,c_n)$ in lieu of $\delta f(X_n,c_n)$ in order to artificially create a sizeable bias.

Under convexity conditions on $f(\cdot)$ and $q(\cdot)$, and under

conditions (which hold in our simulations) on $\{a_n, b_n, c_n, \xi_n\}$, $X_n \to \theta$, the optimal point, w.p.1, as $n \to \infty$ [6]. It was not proved that the multipliers λ_n converged - although they did so in all our simulations, when the $q_i(x)$ were convex in a neighborhood of θ. The simulations were each of length 500; occasionally if it did not appear that λ_n was converging, we took a run of length 25,000. Unless the noise was small ($\sigma^2 \leq \frac{1}{2}$) or $f_x(x)$ + noise observed, the convergence of $\{\lambda_n\}$ was slow - in oscillatory cycles of increasing duration and decreasing variation. When some $q_i(\cdot)$ was not locally convex at θ, the situation was less clear. Sometimes $\{\lambda_n\}$ seemed to converge, and sometimes the variations (max-min) of the cycles did not obviously appear to be diminishing.

In the convergence proof of [6] $a_n = b_n$ was used, but the proof can be readily modified so that $a_n \neq b_n$, provided that both sequences decrease at the same rate. Let $A_\lambda > 0$ and $A_x > 0$, and set $b_n = A_\lambda / m_n^\delta$, $a_n = A_x / m_n^\delta$, $n \leq 100$, and $b_n = A_\lambda / m_n^\alpha$, $a_n = A_x / m_n^\alpha$, $n > 100$, and $c_n = C/(n+1)^\gamma$.

An asymptotic result.[+] Assume that λ_n converges to some $\bar{\lambda}$. If $q_i(\theta) = 0$, assume $\bar{\lambda}^i > 0$. For the asymptotic result we can drop all constraints for which $q_i(\theta) < 0$; thus, suppose that

[+] The subsection cites a result from [10], which has not appeared. The result is not crucial to this paper, but it does help with the understanding of the numerical data.

$q_i(\theta) = 0$, all i. Define $Q = [q_{1,x}(\theta),\ldots,q_{s,x}(\theta)]$, Q_i = Hessian of $q_i(\cdot)$ at θ, F = Hessian of $f(\cdot)$ at θ, $\bar{K}_0 = (F + \sum_i \bar{\lambda}^i Q_i)$, I_t = identity matrix in R^t,

$$\bar{K} = \begin{bmatrix} (F + \sum_i \bar{\lambda}^i Q_i) & Q \\ -Q & 0 \end{bmatrix},$$

$$\bar{K}_2 = -\begin{bmatrix} A_x I_r & 0 \\ 0 & A_\lambda I_s \end{bmatrix} \bar{K}, \quad \bar{K}_1 = \beta I_{r+s} - \bar{K}_2 .$$

Let (in this subsection only) $a_n = A_x/(n+1)^\alpha$, $b_n = A_\lambda/(n+1)^\alpha$, $c_n = C/(n+1)^\gamma$. Let $\mathrm{Cov}[\xi_n \mid \xi_0,\ldots,\xi_{n-1}] \to \sigma^2 I_r$ as $n \to \infty$. Let $\bar{U}^i$, $i = 1,2$, denote random vectors whose distribution is the stationary distribution of the solution to (V^i is s dimensional, U^i is r dimensional)

$$(3.2) \qquad d\begin{pmatrix} U^i \\ V^i \end{pmatrix} = K\begin{pmatrix} U^i \\ V^i \end{pmatrix} - \begin{pmatrix} A_x BC^2 \\ 0 \end{pmatrix} dt - \begin{pmatrix} A_x \sigma/2C \\ 0 \end{pmatrix} dw,$$

where $K = \bar{K}_1$ or $\bar{K}_2$, according to the case. Then, if we normalize the errors as $(n^\beta(X_n - \theta), n^\beta(\lambda_n - \bar{\lambda}))$, the asymptotic result of Section 2 holds [10]; also for the one-sided difference form of (3.1), where $C, \bar{B}$ and A_x/C replace C^2, B and $A_x/2C$, resp.

The asymptotic rate result can conceivably be used to develop some sort of adaptive method for adjustment of the

coefficients A_λ, A_x.

Numerical data. Refer to Figures 3.1 to 3.12. Let

$$q_1(x) = -\sin x_1 + x_2$$

$$q_2(x) = x_1/4 - x_2/2.5 - 1$$

$$q_3(x) = .1(x_1-1)^2 - x_2 - 3 ;$$

(in Figures 3.10-3.12, $q_2(x) = x_1/\pi - x_2/2.5-1$), $f(\cdot)$ is as in Section 2, but may be translated and rotated as indicated in the figures. The values of $A_x, \alpha, \beta, \delta, C$ are as in Section 2, for the same reasons. An exception is that $\delta = \frac{1}{2}$ in Figures 3.2b, 3.4 and 3.8 to 3.12, since we plotted runs taken before the value $\delta = \frac{2}{3}$ was settled upon. When the ellipse ($f(\cdot)$) is centered at $(2\frac{1}{2}, 1\frac{1}{2})$, the constraints are locally convex at[+] θ, but then there is a second local minimum at $\theta_1 = (-3.21, .0683)$ with magnitude 10.2. When the ellipse is centered at $(-2\frac{1}{2}, \frac{1}{2})$ or at $(4,0)$, $q_1(\cdot)$ is not locally convex at θ. Also, the left-hand tail of the constraint set is "thin", which causes some problems when we do not assume knowledge of $q(\cdot)$, but instead take noise corrupted observations of the form $q(X_n)$ + observation noise.

Refer to Figures 3.1 to 3.4 (Figure 3.1 depicts a rather poor run), where the ellipse is centered at $(2\frac{1}{2}, 1\frac{1}{2})$. In (the typical) Figures (3.1,3.2), $A_\lambda = 2$, and $A_\lambda = 5$ in Figure 3.3.

[+] $\theta = (1.85, .961)$, $f(\theta) = .396$.

The multiplier $\bar{\lambda}^1$ at θ is 1.089. We set $\lambda_0 = 0$. Since $\lambda_0 = 0$, the procedure starts off as a pure gradient procedure, and, initially, heads toward the unconstrained minimum. Thus, soon, $q_1(X_n) > 0$. Then λ_n^1 increases, which keeps the path from wandering too far from the constraint set. Progress toward θ is hindered by the local minimum at θ_1, which seems to divert the path until iteration 8 or so. The Lagrangian method was relatively insensitive (compared with the algorithms of Sections 4 and 5) to this local minimum. (The other algorithms usually converged to θ_1, when X_0 was to the left of θ_1.) The asymptotic rate result, and the fact that the Hessian of $q_1(\cdot)$ is not positive definite at θ_1, suggest that the non-convex curvature of $q_1(\cdot)$ at θ_1 makes the iterate sequence less stable in the vicinity of θ_1.

As λ_n^1 increases, the path moves back to the feasible set, and continues to leave and enter the set, following the oscillations of λ_n^1, which in turn are caused by the variations in the values of the $q_1(X_n)$ sequence. As λ_n^1 increases, the effect of $\lambda_n^1 q_{1,x}(X_n)$ on the $\{X_n\}$ increases, helping to push the iterates to the constraint set (i.e., to reduce the value of $q_1(x)$). Once X_n is inside the constraint set, λ_n decreases, and the effect of the gradient $f_x(X_n)$ becomes more pronounced. All this is as expected.

We used $M_\lambda^i = 2$. The value of this upper bound is quite important. If too large, the oscillations of $\{\lambda_n\}$, and consequently those of $\{X_n\}$, in the initial phases, are larger - and the graphs have a substantially wilder appearance. If M_λ^i

is too small, then the limit point may not be feasible. Perhaps it is best to begin with a conservatively small value - and to increase it if the sequence seems to be converging to a non-feasible point - or if the path $\{X_n\}$ is well behaved.

Observe the effect of A_λ (Figures 3.1, 3.3). With the larger value, the oscillations of $\{\lambda_n\}$ are slightly more pronounced. With $A_\lambda = 5$, the results were preferable to those for $A_\lambda = 2$. As n increased, the period and the magnitude of the $\{\lambda_n\}$ fluctuations decreased in all cases. With $A_\lambda = 1$ (not plotted), there were only two obvious oscillations in the first 500 iterations; they were fairly smooth, and the tail of the second oscillation appeared to be settling near some "steady state" value. Possibly, more insight into how to select A_x, A_λ can be obtained from a study of the asymptotic result.

A number of variations on the method of adjusting $\{\lambda_n\}$ were tried (in the absence of convergence proofs). In some runs, we let A_λ be a diagonal matrix, whose i^{th} element (at n) depended on the sign of $q_i(X_n)$, generally being larger for $q_i(X_n) < 0$ than for $q_i(X_n) > 0$. The results were not particularly good, but the method was motivated by a desire to force λ_n to zero as rapidly as possible when the iterates were inside the feasible region and, hence, to decrease the magnitude of the path oscillations when in that region. A method that worked somewhat better used sign $q_i(X_n)$ in place of $q_i(X_n)$ in (3.1a). A typical run is plotted in Figure 3.2. The oscillations of $\{\lambda_n\}$ were

larger - and it was not clear whether $\{\lambda_n\}$ ever converged. But $\{X_n\}$ did appear to converge - and often did better than for the original scheme (3.1). Although the theoretical properties of this modification are not known, it is suspected that some such variation of the λ_n iteration formula would yield more stable $\{X_n\}$ paths, and nice convergence.

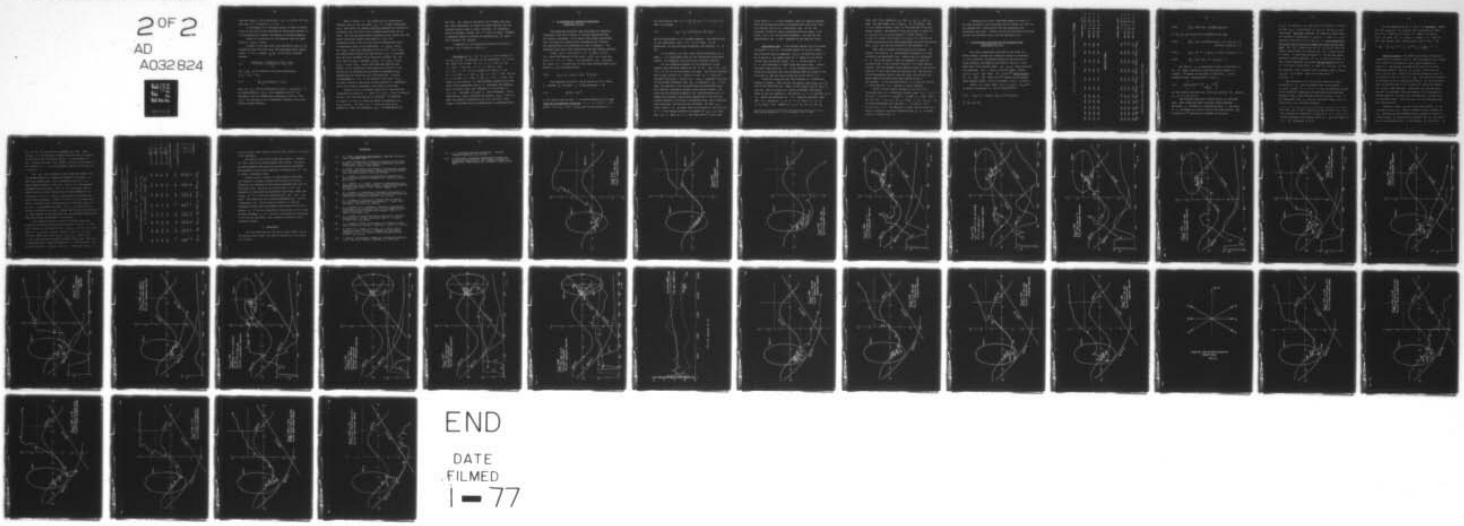
If we assumed that the observation was $f_x(X_n)$ + noise, instead of $f(X_n)$ + noise, then the results were extremely good; there were only small oscillations of the $\{X_n\}$ path as it tended to θ. The path was (roughly) a smooth version of those in the figures until iteration 20 or so, after which it moved to θ roughly along the upper boundary, with only small oscillations. Also, $\{\lambda_n\}$ settled down quite fast. See Figure 3.8, a typical run for the value $\sigma^2 = \frac{1}{2}$.

Now refer to Figures 3.5 to 3.8, where $q_1(\cdot)$ is not locally convex at θ, and the multiplier $\bar{\lambda}^{-1}$ at θ is 0.464. Also, $\theta = (-2.7,-.427)$, $f(\theta) = .255$. The sequence $\{X_n\}$ (experimentally) converged, despite the lack of convexity. In Figures 3.5 and 3.6, $\sigma^2 = \frac{1}{2}$. This "non-convex" case seemed to be more sensitive to noise - but, did well for $\sigma^2 = \frac{1}{2}$. Figure 3.7 illustrates the increased oscillations when $\sigma^2 = 2$ (although the plotted run was a little better than average). In Figures 3.5 to 3.7, it appears that the λ_n are settling near zero. Their behavior for $n \leq 25{,}000$ is similar to that in Figure 3.13.c, which is for the data of Figure 3.6, but with $\sigma^2 = 2$.

AD-A032 824 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
I, II CONVERGENCE AND RATE OF CONVERGENCE THEOREMS FOR CONSTRAI--ETC(U)
AUG 76 H J KUSHNER, S LAKSHMIVARAHAN N00014-76-C-0279

UNCLASSIFIED LCDS-TR-76-1 NL

2 OF 2
AD
A032824



END

DATE
FILMED
1-77

Note also Figure 3.13a,b, which plot $\{\lambda_n\}$, $n \leq 25{,}000$, for cases with the $f(\cdot)$ of Figures 3.1 to 3.4.

In the absence of a convergence proof, we make no general claim for the value of the algorithm for non-convex problems, but it is interesting to observe its behavior on selected examples.

Figure 3.8 illustrates the fine behavior when the observation is $f_x(X_n)$ + noise.

A number of runs were taken with observation noise on the constraints also (for which case, the convergence proof remains valid). In (3.1), replace $q_{x_i}(X_n)$ by the noise corrupted estimate

$$\frac{q_i(X_n+e_ic_n) - q_i(X_n-e_ic_n) + \psi^+_{i,n} - \psi^-_{i,n}}{2c_n} \tag{3.3}$$

and $q_i(X_n)$ by the average of the noisy observations $q_i(X_n \pm e_ic_n) + \psi^{\pm}_{i,n}$

$$\frac{1}{2r} \sum_{i,\pm} [q_i(X_n \pm e_ic_n) + \psi^{\pm}_{i,n}], \tag{3.4}$$

where the $\psi^{\pm}_{i,n}$ are the observation noises on $q_i(X_n \pm e_ic_n)$. In the simulations, the $\{\psi^{\pm}_{i,n}\}$ were independent of the $\{\xi_n\}$ (although it was not theoretically necessary to do so). Also $\{\psi^{\pm}_{i,n}\}$ was selected to be an independent sequence of zero mean Gaussian random variables.

Refer to Figure 3.9. The results are not significantly different than for the case where $\psi^{\pm}_{i,n} \equiv 0$, a rather encouraging sign - since there are numerous practical problems with constraints of unknown form, but where noise corrupted observations can be taken. The data are particularly interesting since (apart from some rather inefficient penalty algorithms [3], [14]), the Lagrangian method is the only one known to converge when there is observation noise on the constraints. If $q_x(X_n)$ + noise and $q(X_n)$ + noise were observed in lieu of (3.3), (3.4), then the results were virtually the same as with $\psi^{\pm}_{i,n} = 0$, provided that the noise corrupted finite difference estimate $\delta\overline{Y}_n$ was still used in order to estimate $f_x(X_n)$. If $f_x(X_n)$ + noise, $q_x(X_n)$ + noise, $q(X_n)$ + noise, were observed, then the main effects of the "constraint" noise seemed to be similar to the effects due to increasing the performance measurement noise somewhat, and the iterations still converged well. Owing to the dependence on $\{\lambda_n\}$ of the effects of the noise $\{\psi^{\pm}_{i,n}\}$, the exact effects of this noise are more difficult to understand, and we do not have a convergence rate result in this case.

Figures 3.10 to 3.12 depict results for a case where two constraints are active at θ. Again, if $f_x(X_n)$ + noise is observed (Figure 3.10), the path behavior is excellent. A comparison of Figures 3.11 and 3.12 shows the effect of quadrupling the noise variance. The value $M^i_\lambda = 2$ which was used may be slightly too small for constraint $q_2(\cdot)$, as suggested by the

fact that $\{\lambda_n^2\}$ stays at the value 2 for a rather long time. But, since convergence (at least to a point very near the constrained minimum) seems to occur, and since λ_n^2 eventually starts to decrease, perhaps $M_\lambda^2 = 2$ is not too small. Probably, a reasonable adaptive procedure for adjusting the M_λ^i would have increased M_λ^2 slightly.

A comparison of the Lagrangian algorithm with those of Sections 4 and 5 appears in Table 5.1.

The effect of A_λ. Let $f(\cdot), q(\cdot)$ be as in Figure 3.1, and let $A_x = \frac{1}{4}$. A calculation of the eigenvalues (ρ) of the $-\bar{K}_2$ above (3.2) yields: for $A_\lambda = 1$, $\rho_1 = .396$, ρ_2 and $\rho_3 = .245 \pm i\ .459$; for $A_\lambda = 5$, $\rho_1 = .395$, ρ_2 and $\rho_3 = .245 \pm i\ 1.39$. For the situation of Figure 3.5, for $A_\lambda = 1$, we have $\rho_1 = .322$, ρ_2 and $\rho_3 = .127 \pm i\ .643$ and for $A_\lambda = 5$, we have $\rho_1 = .308$, ρ_2 and $\rho_3 = .134 \pm i\ 1.493$. For fixed A_x, the value of A_λ had negligible effects on the real parts of the eigenvalues, but as A_λ increased, the imaginary parts increased; hence, the "frequency" of oscillation of $\{\lambda_n\}$, $\{X_n\}$ increased, for large n. This "frequency", result holds because, under a suitable time scaling, $\{X_n, \lambda_n\}$ have the same "oscillatory" properties as do (3.2); see [10]. This remark is consistent with our observations, as discussed above.

4. An Algorithm for Inequality Constraints (Algorithm 3 of [8])

The augumented multiplier type algorithms for inequality constraints (here and in Section 5) are more complicated than the previous algorithms. On the whole, they worked surprisingly well, even when the conditions for convergence were violated. The algorithm here is an adaptation of that of Section 2. All undefined terms retain their definitions from the previous sections.

Define $\tilde{q}_i(x) = \max[0, q_i(x)]$, $P(x) = \sum_i \tilde{q}_i^2(x)$, and let $\Phi(x)$ denote the Jacobian of the vector $\{q_i(x)$, all i such that $q_i(x) \geq 0\}$. Let $\Phi_n = \Phi(X_n)$. The constraints $q_i(\cdot)$ are known, but, again, $f(\cdot)$ is known only via the noise corrupted observations. Define (λ_n will be defined below)

$$X_{n+1} = X_n - a_n[\delta Y_n + \Phi_n'\lambda_n + \frac{k}{2} P_x(X_n)]. \tag{4.1}$$

The algorithm of Section 2 can be written as (4.1), where ϕ_i replaces $\tilde{q}_i$ and where λ_n is the minimizing λ in

$$\min_{\lambda} |\delta Y_n + \Phi_n'\lambda|^2. \tag{4.2}$$

In the current case, we take λ to be the minimizer in (4.2) <u>but under the non-negativity constraint</u> $\lambda^i \geq 0$ ($\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^s)'$). Define $\pi^+(x)$ by: $\pi^+(x)v$ is the error in the projection of v

onto the positive cone $\{y: y = -\sum_i \lambda^i q_{i,x}(x),\ \lambda^i \geq 0,\ q_i(x) \geq 0\}$. Then (4.1) becomes

$$X_{n+1} = X_n - a_n[\pi^+(X_n)\delta Y_n + \frac{k}{2} P_x(X_n)], \tag{4.3}$$

and the relationship to (2.1) becomes apparent. The analysis here is more complicated, since $\pi^+(x)$ is not linear. In our simulations, we used one-sided differences, and replaced δY_n by $\delta \overline{Y}_n$.

It is proved in [8] that, under various conditions, $X_n \to \theta$ where θ is a Kuhn-Tucker point for the constrained optimization problem. No asymptotic rate results are currently available.

For our simulated problems, all the conditions for convergence were satisfied, except for (A6) of [8]. (The $\tilde{\pi}^+(x)$ and $\pi^+(x)$ of (A6) of [8] are actually the same.) The finite difference form of this condition is: there is a real $k_1 > 0$ such that $f_x'(X_n)E[\pi^+(X_n)(f_x(X_n) + \xi_n/2c_n) \mid X_0,\ldots,X_{n-1}] \geq k_1 f_x'(X_n)\pi^+(X_n)f_x(X_n)$. (In the forward difference case $2c_n$ is replaced by c_n.) In our case, the condition is not satisfied near the Kuhn-Tucker point. Yet, numerical convergence seemed to have always occurred. The point will be developed elsewhere, but, from preliminary work, it appears that suitable (non-obvious) adaptations of the weak convergence methods in [9] yield a convergence proof, under conditions which hold here. We note only the following: for each c_n, a vector field on R^r is drawn, where the direction of the flow line at x is the average direction of movement of $X_{n+1} - X_n$, given that $X_n = x$. Then, as $n \to \infty$, the flows tend to a flow, whose

lines "flow" to θ. On the boundary, there is a special problem, since the average directions are discontinuous, and in the limit, the "average" flow is of a "chattering" type. So far, the technique of proof of actual convergence does not appear to be simple, and it will be developed elsewhere. The need for the generalization was, in fact, suggested by the numerical study.

Experimental data. A few selected typical runs are plotted in Figures[+] 4.1 to 4.4. The method appears to be quite robust. Apart from the usual SA sequences $\{a_n, c_n\}$, we need only select k. As in Section 2, the larger values of k appear to give better results than the smaller values. For k in the range of .5 to 4, the differences in final values were quite small. Several runs were taken using the $f(\cdot)$ of the Lagrangian Figure 3.1. Unlike the Lagrangian case, this procedure often got caught at the local min θ_1, when the initial point was to the left of θ_1.

Inside the constraint set, the procedure is a classical SA gradient procedure. When outside, the penalty terms $P_x(X_n)$ pull the path to the constraint set. There is also another type of "pull" to this set. Refer to Figure 4.5, where we consider a case where there is only one constraint. The vectors v_i, $i \leq 4$, represent $\delta\bar{Y}_n$ or δY_n. Recall the definition of $\pi(x)$ (equality constraint case). Then $\pi(x)$ applied to v_1 or v_3 is v_5 and to v_2 or v_4 is v_6, while $\pi^+(x)$ applied to v_3 or v_4 are v_5 and v_6,

---

[+]Here and in Section 5, $\theta = (-2.7, -0.427)$, $f(\theta) = .255$.

resp., and $\pi^+(x)$ applied to v_1 and v_2 are v_1 and v_2, resp. For this reason, the $-\pi^+(X_n)\delta Y_n$ (or $-\pi^+(X_n)\delta \bar{Y}_n$) term in (4.3) tends to create a "pull" in the direction of the constraint surface. The effects of this can be seen in Figure 4.3 at iterates 8 to 11, and in Figure 4.4, after iterate 15. In fact, at iterate 15 in Figure 4.3, $f_x(X_n) \approx 0$, and both the biased projections of the random observation noise and the penalty term $P_x(X_n)$ pull the path directly to the constraint set.

Many types of adaptive modifications are possible. For example, the value of a_n could depend on whether or not some constraint was satisfied or not. Observe that the variations in X_n tend to be greater (when the sequence is inside the constraint set) in the vertical direction - for rather obvious reasons. Once the sequence values improve enough in the vertical direction so that they leave the constraint set, then the projection and penalty effects come into play, and the path then moves toward the optimal point while staying close to the boundary. In the Lagrangian method, it is only the "prices" λ_n, which keep the $\{X_n\}$ feasible or nearly feasible.

Table 4.1 gives sample averages for 20 runs of 500 iterations each on a 5 dimensional problem, where $f(x) = (x_1-1)^2 + .5(x_2-2)^2 + .8(x_3-1)^2 + .4(x_4+\frac{1}{2})^2 + .2(x_5+\frac{1}{2})^2$, $q_1 = x_1^2/4 + x_2^2/2 + x_3^2 + x_4^2 + x_5^2 - 1$, $q_2 = x_2 - x_1^2$, $q_3 = 4x_3x_4 - 1$, $q_4 = -x_4$, $q_5 = -x_5$. one-sided differences were used and $\theta = (1,0,\frac{1}{2},0,0)$, $q(\theta) = (0,0,-1,0,0)$, $k = 4$. The behavior for other initial points was similar, generally improving (becoming worse) as X_0 was moved closer to (further from) θ.

A comparison with other algorithms appears in Table 5.1. The runs analogous to the Lagrangian runs in Figures 3.11 and 3.12 also behaved well, indeed better than the Lagrangian runs - convergence was faster. Similar runs for the algorithm of the next section also performed better than the Lagrangian runs.

5. An Alternative Algorithm for the Inequality Case (Algorithm 4 of [8])

This (more complicated) algorithm is the result of a direct attempt to handle the inequality problem by converting it into an equality constrained problem via the addition of slack variables, and adapting the method of Section 2, from which some of the notation comes. Let $\{w_n, v_n\}$ denote positive null sequences. Let b denote a positive number. Define $\phi_i(x,z) = q_i(x) + bz^i$, where bz^i is the i^{th} slack variable. Let $Z_n = (z_n^1, \ldots, z_n^s)$ denote the n^{th} estimate of the optimal slack variable vector, $-q(\theta)/b$, and define $P(x,z) = \sum_i \phi_i^2(x,z)$.

Define $\Phi_n = \Phi(X_n)$, where $\Phi(x) = [\phi_{1,x}(x), \ldots, \phi_{s,x}(x)]'$, as earlier, and define (λ_n will be defined below)

$$(5.1) \qquad X_{n+1} = X_n - a_n[\delta Y_n + \Phi_n'\lambda_n + k/2\ P_x(X_n, Z_n)].$$

If $z_n^i > v_n$, set

mean values of 20 runs, each of 500 iterations

	$\bar{f}(X_{500})$	$\bar{q}_1(X_{500})$	$\bar{q}_2(X_{500})$	$\bar{q}_3(X_{500})$	$\bar{q}_4(X_{500})$	$\bar{q}_5(X_{500})$	$\bar{x}^1_{500}$	$\bar{x}^2_{500}$	$\bar{x}^3_{500}$	$\bar{x}^4_{500}$	$\bar{x}^5_{500}$
1. $A = \frac{1}{4}$, $\sigma^2 = \frac{1}{2}$, $\delta\bar{Y}_n$ used	.934	-.021	-.046	-.964	-.021	-.069	1.011	.975	.472	.020	.068
2. $A = \frac{1}{8}$, $\sigma^2 = \frac{1}{2}$, $\delta\bar{Y}_n$ used	1.00	-.022	-.026	-.965	-.013	-.051	.911	.806	.649	.012	.050
3. $A = \frac{1}{8}$, $\sigma^2 = 2$, $\delta\bar{Y}_n$ used	1.21	-.092	-.104	-.879	-.057	-.138	.915	.751	.562	.050	.138
4. $A = \frac{1}{4}$, $\sigma^2 = 1$, $f_x(x)$ + noise observed	.850	.003	-.001	-.998	-.001	-.006	1.00	.100	.499	.001	.005

Sample Variances

1. $A = \frac{1}{4}$, $\sigma^2 = \frac{1}{2}$, $\delta\bar{Y}_n$ used	.005	.002	.008	.005	.001	.005	.003	.006	.008	.002	.003
2. $A = \frac{1}{8}$, $\sigma^2 = \frac{1}{2}$, $\delta\bar{Y}_n$ used	.0182	.002	.002	.004	.001	.003	.004	.015	.024	.006	.014
3. $A = \frac{1}{8}$, $\sigma^2 = 2$, $\delta\bar{Y}_n$ used	.107	.012	.015	.035	.005	.016	.016	.056	.024	.006	.014
4. $A = \frac{1}{4}$, $\sigma^2 = 1$, $f_x(x)$ + noise observed	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000

TABLE 4.1. A 5 Dimensional Problem. Initial Point (-3,-3,-3,-3,-3).

(5.2a) $$z_{n+1}^i = \max[0, z_n^i - a_n(b\lambda_n^i + k\phi_i(X_n, Z_n))].$$

If $z_n^i \leq v_n$, use (5.2b,c or d) according to the case.

(5.2b) $$z_{n+1}^i = z_n^i - a_n(b\lambda_n^i + k\phi_i(X_n, Z_n)), \text{ if } \phi_i(X_n, Z_n) \leq 0 \text{ and } b\lambda_n^i + k\phi_i(X_n, Z_n) \leq 0,$$

(5.2c) $$z_{n+1}^i = z_n^i \text{ if } \phi_i(X_n, Z_n) \leq 0, b\lambda_n^i + k\phi_i(X_n, Z_n) > 0,$$

(5.2d) $$z_{n+1}^i = z_n^i + a_n w_n \text{ if } \phi_i(X_n, Z_n) > 0.$$

The value of λ_n is selected by a type of projection [8] (in x,z space), similarly to what was done in Section 2 in x space). The method is equivalent to selecting λ_n to be a minimizing vector (unconstrained in sign) in

(5.3) $$\min_{\lambda} \left[|\delta Y_n + \Phi_n' \lambda|^2 + b^2 \sum_{i: z_n^i > v_n} (\lambda^i)^2 \right]$$

(in the forward difference case, which was simulated, $\delta \bar{Y}_n$ replaces δY_n in (5.1) and (5.3)).

The algorithm seems more involved than it is, and works well. Under conditions which hold here (unless otherwise mentioned) X_n converges to a Kuhn-Tucker point θ. The value of v_n determines an a priori level of the slack variable bz_n^i at which the i^{th} constraint is presumed to be active.

If $q_i(\cdot)$ is inactive at X_n, then z^i_{n+1} is calculated in the same way that X_{n+1} is calculated.[+] If $z^i_n \leq v_n$, the rule is more complicated. Note that (although, for technical reasons which can probably be circumvented, $\sum (a_n/v_n)^{\ell} < \infty$, some $\ell > 0$, was required in the proof in [8]) if $v_n \equiv 0$ in our simulations, the results were as good as or better than those for the other cases. The $\{w_n\}$ sequence and the λ^i_n in (5.2a,b) serve the purpose (among others) of assuring that $\{X_n\}$ cannot converge to a point $\bar{\theta}$ where some multiplier $\bar{\lambda}$ (in the Kuhn-Tucker condition) would be negative. If X_n is near such a point $\bar{\theta}$, then (loosely speaking) the general average negative value of λ^i_n (for large n) would force X_n away (if $\phi_i(X_n,Z_n) \leq 0$) and the iteration (5.2d) (together with the penalty effect, forcing $\phi_i(X_n,Z_n)$ to be small) would do the same thing if $\phi_i(X_n,Z_n) > 0$.

The calculation (5.3) can also be viewed as a type of projection of δY_n onto the tangent plane to the surface $\{y: q_i(y) - q_i(X_n) = 0, i \leq s\}$ at $y = X_n$, but where the weights of the "inactive" constraints are penalized by the additional quadratic term, so that they will not be too large. If all λ^i are "penalized", then λ_n will be "relatively" small, and (5.1) behaves more like a gradient procedure.

[+] Note that $\lambda^i_n b$ is just $\lambda^i_n \partial\phi_i(X_n,Z_n)/\partial z^i_n$, and that $\delta Y_n + \Phi'_n\lambda_n$ is an estimate of $\text{grad}_x[f(x) + \sum_i \lambda^i_n\phi_i(x,z)]$ at $x = X_n$, $z = Z_n$. The two expressions are similar; since $f(\cdot)$ does not depend on z, no δY_n need appear in (5.2).

In our simulations, we let k and b be variables. Replace $b\lambda_n^i$ and $k\phi_i(X_n,Z_n)$ by $b_i(X_n)\lambda_n^i$ and $k_i(X_n)\phi(X_n,Z_n)$, where $b_i(x) = b_1$ (resp. b_2) if $q_i(X_n) < 0$ (resp., ≥ 0), and set $k_i(x) = k_1$ (resp., k_2) if $q_i(X_n) < 0$ (resp., ≥ 0). Also, we used $\delta\bar{Y}_n$ and set $bZ_0^i = \max[0,-q_i(X_0)]$, and $A = \frac{1}{4}$, $\alpha = \frac{5}{6}$

$$\delta = \frac{2}{3},\ \gamma = \frac{1}{6},\ w_n = n^{-.4},\ v_n = Vn^{-1/6},\ c_n = \frac{1}{2n^{1/6}},\ \sigma^2 = 2.$$

Numerical results. The general qualitative features are illustrated in Figures 5.1 to 5.6. Note the behavior from iterate 20 to iterate 80 in Figure 5.1, and a similar behavior in Figure 5.2. The looping is caused by the bias in the finite difference estimate $\delta\bar{f}(X_n,c_n)$ of $f_x(X_n)$. Recall that our bias is quite large. *Here, it is just big enough at iterate 20* to reverse the sign of $E[\lambda_n^1 | X_n \approx X_{20}]$ from positive to negative. So the forms (5.2a, b) increase the value of Z_n^1 on the average, hence the penalty term in (5.1) forces X_n^1 inside the feasible set ($q_1(X_n)$ must decrease). Eventually, the bias decreases. The point is important, since the behavior of an algorithm can depend heavily on the form of the available data, as well as on the algorithm itself. With less bias, all the algorithms would have behaved better.

Generally, larger w_n does better than smaller w_n. As indicated earlier, $\{w_n\}$ plays an important role in pushing the $\{X_n\}$ sequence away from boundary points (toward the interior). If the average λ_n^i values are all positive, the X_n will drift back to the boundary point. If the average λ_n^i are not all ≥ 0,

the w_n and λ_n^i play the role of pushing X_n^i away. (When $\phi_i(X_n,Z_n) > 0$, for "penalty" reasons, we do not want to increase Z_n, but we do want to decrease $q_i(X_n)$.) As the elements of $\{w_n\}$ increase, the path $\{X_n\}$ tends to be more on the inside of the boundary (if θ is on the boundary) than on the outside, as it moves to θ.

The $\{X_n\}$ has a tendency to move along the boundary (see the typical Figure 5.6), owing to the method of adjusting Z_n, and to the penalty term. This is especially true if k is large (compare Figure 5.2 with 5.6). With $k = 1$, the procedure converged much more slowly. But, if k is too large, then it appears that the greater penalty puts too much emphasis on the convergence to zero of $P(X_n,Z_n)$, which seems to slow down the actual path convergence. Figures 5.3 and 5.4, for the same parameters, illustrate great differences due to different noise. If $b_1 = b_2 = 1$, convergence was slow. If $b_1 > b_2$, then if a constraint is satisfied, it plays a smaller role in the calculations of λ_n, and, interior to the constraint set, the procedure behaves more like the gradient SA procedure, as it should.

Now, compare Figures 5.1 and 5.4, the data is the same except for the noise sequence and the value of (k_1,k_2). In the run of Figure 5.4, the signs of the $\phi_i(X_n,Z_n)$, $n = 2,3$ were such that the $\phi_{i,x}(X_n,Z_n)\cdot\phi_i(X_n,Z_n)$ terms, $i = 2,3$, in $P_x(X_n,Z_n)$ partially offset the effect of the $\phi_{1,x}(X_n,Z_n)$ terms, until $n = 80$. Thus, the path remained "outside" rather long. This type of phenomenon was a common occurrence if k_1 was too small - for then the values of $\phi_i(X_n,Z_n)$ (for the i such that $q_i(X_n)$ 0)

	$\bar{f}(X_{500})$	var $f(X_{500})$	$\bar{q}_1(X_{500})$	var $q_1(X_{500})$	$\bar{x}^1_{500}$	var x^1_{500}	$\bar{x}^2_{500}$	var x^2_{500}
Lagrangian								
1. $A_\lambda = 2$	.415	.045	-.188	n.a.	-2.75	.045	-.56	n.a.
2. $A_\lambda = 5$	.406	.031	-.182	n.a.	-2.76	.042	-.54	n.a.
3. $A_\lambda = 10$	.393	.031	-.162	.056	-2.76	.041	-.522	.088
4. *	.405	.030	-.174	.050	-2.74	.046	-.552	.094
Algorithm of Section 4								
k = 4	.354	.008	-.094	n.a.	-2.75	.047	-.457	n.a.
Algorithm of Section 5**								
1. k = (8,8), b = (1,4), V = 1/16	.349	.015	-.056	.011	-2.71	.058	-.463	.065
2. k = (8,8), b = (1,4), V = 0	.345	.008	-.054	.017	-2.72	.056	-.450	.058
3. k = (8,8), b = (1,2), V = 1/8	.350	.012	-.064	.021	-2.75	.052	-.440	.073
4. k = (2,2), b = (1,4), V = 1/16	.336	.011	-.007	.015	-2.78	.058	-.359	.078

$^*\lambda^i_{n+1} = \max[0, \lambda^i_n + (A_\lambda / m^\alpha_n) \operatorname{sign} q_i(X_n)]$, $\lambda^i_n \leq 2$.

$^{**}v_n = V/n^{.4}$, $w_n = n^{-1/6}$. n.a. = not available, due to omission

$\theta = (-2.7, -.427)$, $f(\theta) = .255$, $A_x = A = 1/4$

TABLE 5.1

A Comparison of Several Algorithms for the Inequality Constrained Case.
Initial Point = (4,3), $\sigma^2 = 2$.
Sample Averages of 20 Runs, each of 500 iterations.

were not always close enough to zero for their effect on $P_x(X_n, Z_n)$ to be negligible.

Two paths in the plotted graphs were typical - although the final values were slightly better than average. The paths are better behaved than those for the previous two algorithms. The procedure is not overly sensitive to variations in b,k and A, within a reasonable range.

Finally, refer to Table 5.1, where several algorithms are compared, via sample averages and variances of 20 runs. Generally, the algorithm of this section performed best, and the Lagrangian worst. The data for initial point (3,-3) are comparable in magnitudes and qualitative properties, except that $\bar{f}(X_{500})$ of 1,2 of the last algorithm were .386 and .373, resp.. The constraints are satisfied reasonably well. For the configuration and initial point of Figure 3.1, only the Lagrangian avoided (and consistently) being caught by the local maximum θ_1. If the last algorithm of this section is to be used, we would recommend $v_n \equiv 0$. The table illustrates the robustness of the various algorithms, under parameter variations (within reasonable ranges).

6. Conclusions

The four algorithms all work well, within certain ranges. The equality case method, and those of Sections 4, 5 are robust and reliable.

REFERENCES

[1] M.T. Wasan, Stochastic Approximation, Cambridge University Press, Cambridge, 1969.

[2] L. Ljung, "Convergence of Recursive Stochastic Algorithms", Report 7403, 1974, Lund Institute of Technology, Division of Automatic Control, Lund, Sweden.

[3] V. Fabian, "Stochastic Approximation of Constrained Minima", Proc. 4th Prague Conference on Statistical Decision Theory and Information Theory", 1966, pp. 277-289.

[4] H. J. Kushner, "Stochastic Approximation Algorithms for Constrained Optimization Problems", Ann. Statist., 2(1974), pp. 713-723.

[5] H. J. Kushner, H. T. Gavin, "Stochastic Approximation-like Algorithms for Constrained Systems: Algorithms and Numerical Results", IEEE Trans. on Aut. Control, AC-19, 1971, pp. 349-357.

[6] H. J. Kushner, E. Sanvicente, "Stochastic Approximation for Constrained Systems with Observation Noise on the System and Constraints", Automatica, 11(1975), pp. 375-380.

[7] H. J. Kushner, E. Sanvicente, "Penalty Function Methods for Constrained Stochastic Approximation", J. Math. Anal. and Applic., 46(1974), pp. 499-512.

[8] H. J. Kushner, M. L. Kelmanson, "Stochastic Approximation Algorithms of the Multiplier Type for the Sequential Monte Carlo Optimization of Constrained Systems", SIAM J. on Control, 14, August 1976.

[9] H. J. Kushner, "General Convergence Results for Stochastic Approximations via Weak Convergence Theory", to appear J. Math. Anal. and Applic.

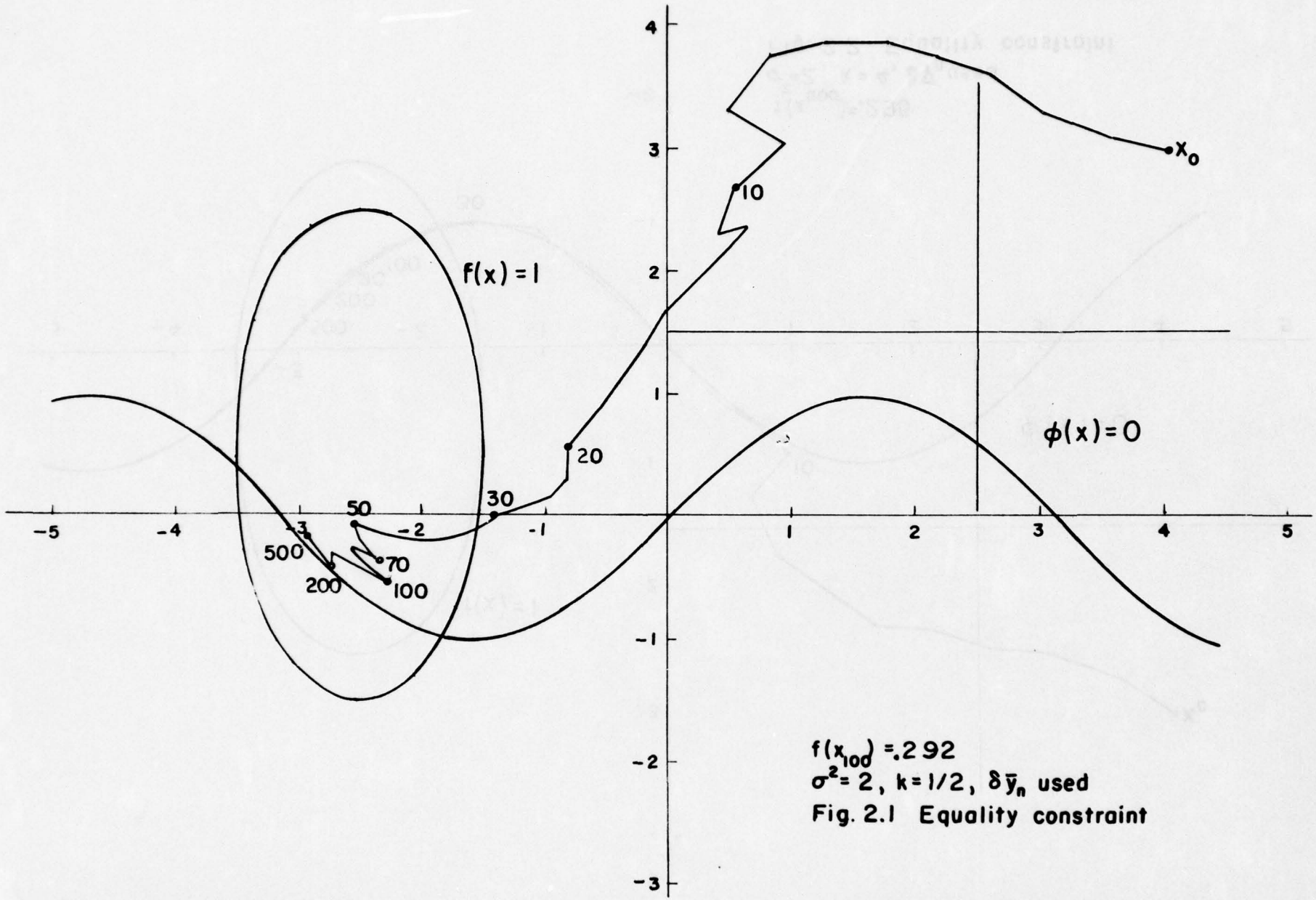
[10] H. J. Kushner, "Rates of Convergence for Sequential Monte Carlo Optimization Methods", submitted to SIAM J. on Control.

[11] A. Miele, E. G. Cragg, R. R. Iyer, A. V. Levy, "Use of Augmented Penalty Function in Mathematical Programming Problems", Part I; J. Optimiz. Theory and Applications, 8(1971), pp. 115-130.

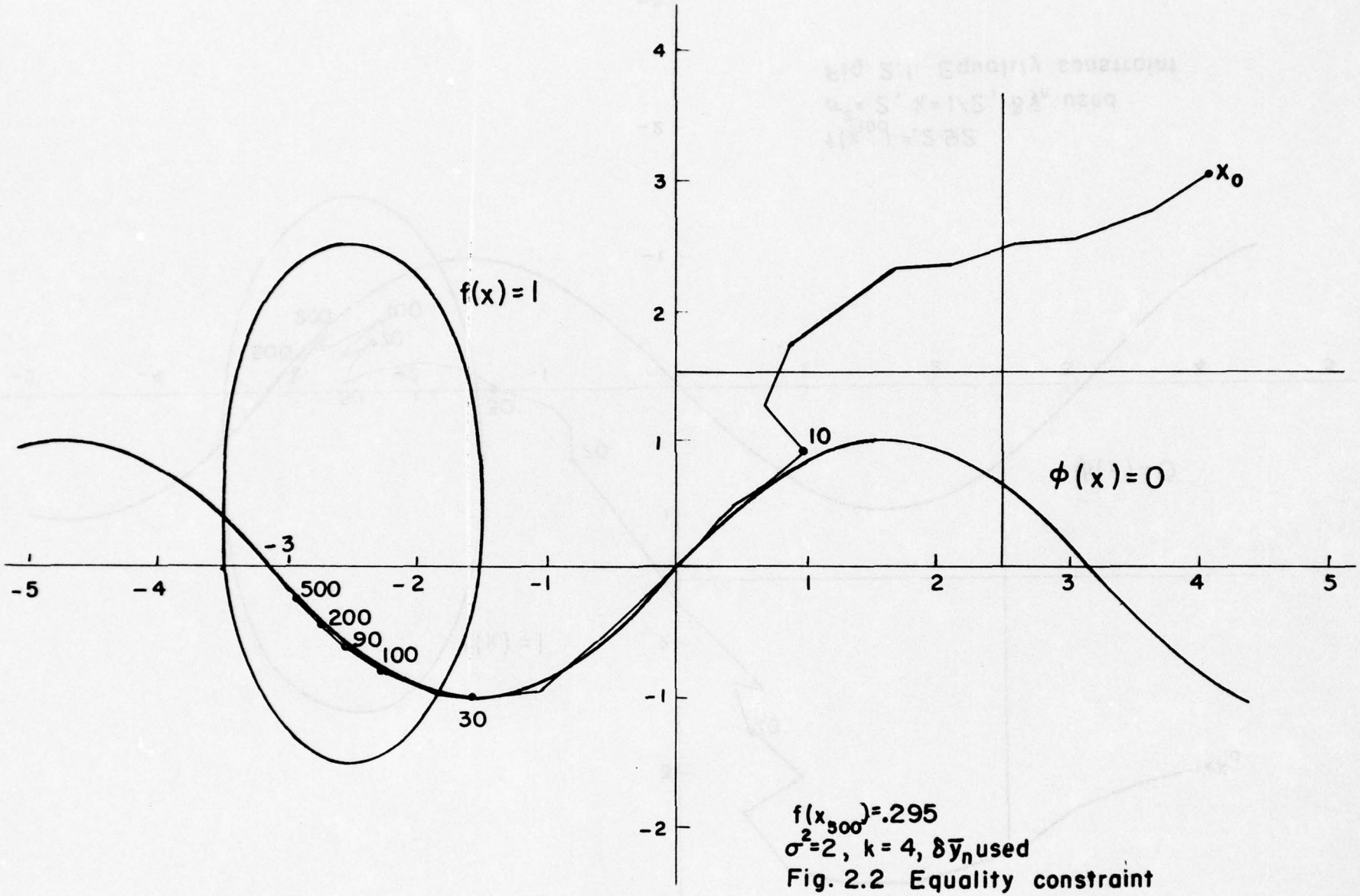
[12] V. Fabian, "On Asymptotic Normality in Stochastic Approximation", Ann. Math. Statist., 39(1968), pp. 1327-1332.

[13] D. I. Bertsekas, "Multiplier Methods: A Survey", Automatica 12(1976), pp. 133-145.

[14] E. Sanvicente, "Stochastic Approximation Methods for Constrained Optimization", Ph.D. Thesis, Division of Engineering, Brown University, Providence, Rhode Island, 1974.

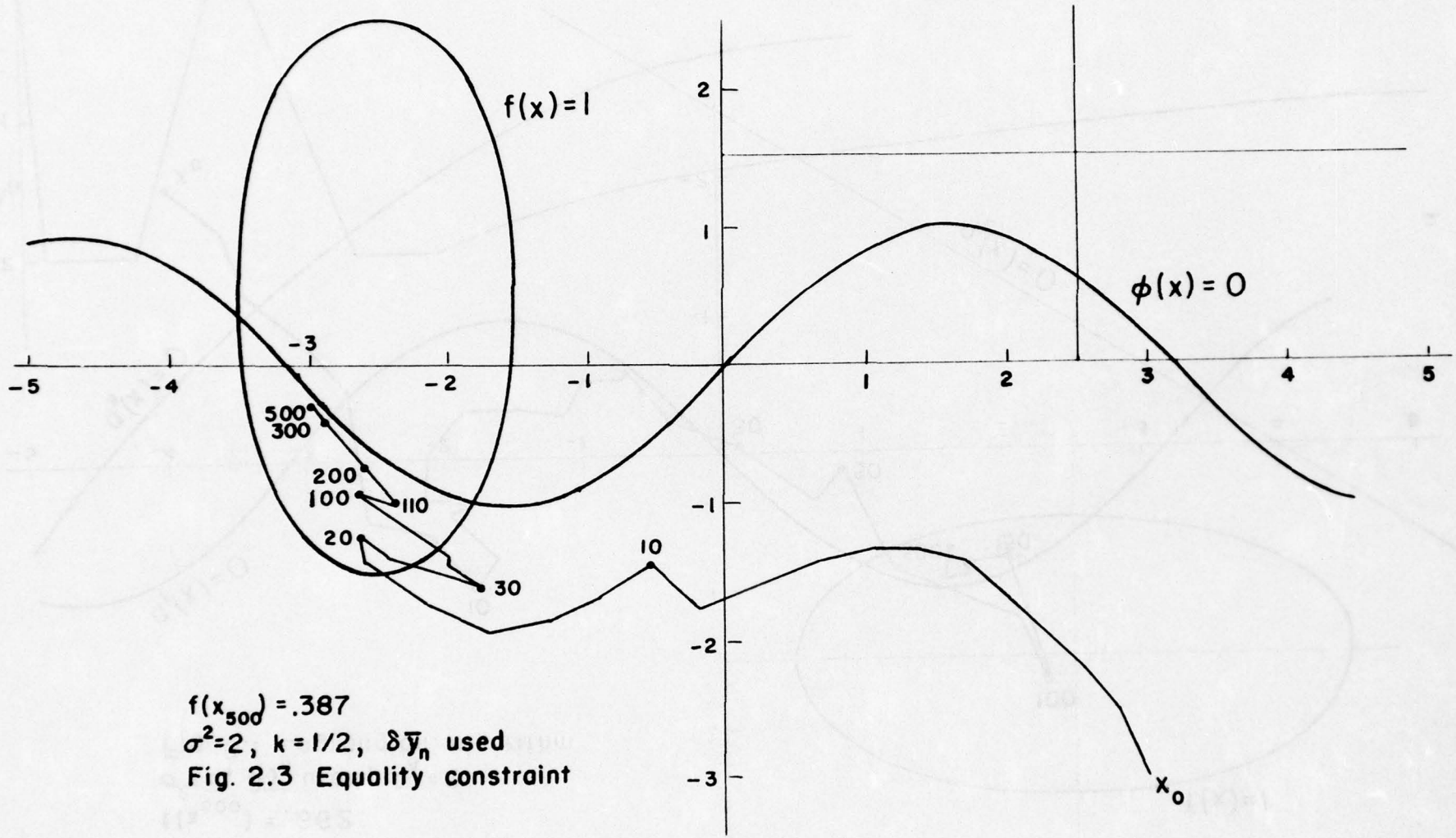


$f(x_{100}) = .292$
$\sigma^2 = 2$, $k = 1/2$, $\delta \bar{y}_n$ used
Fig. 2.1 Equality constraint

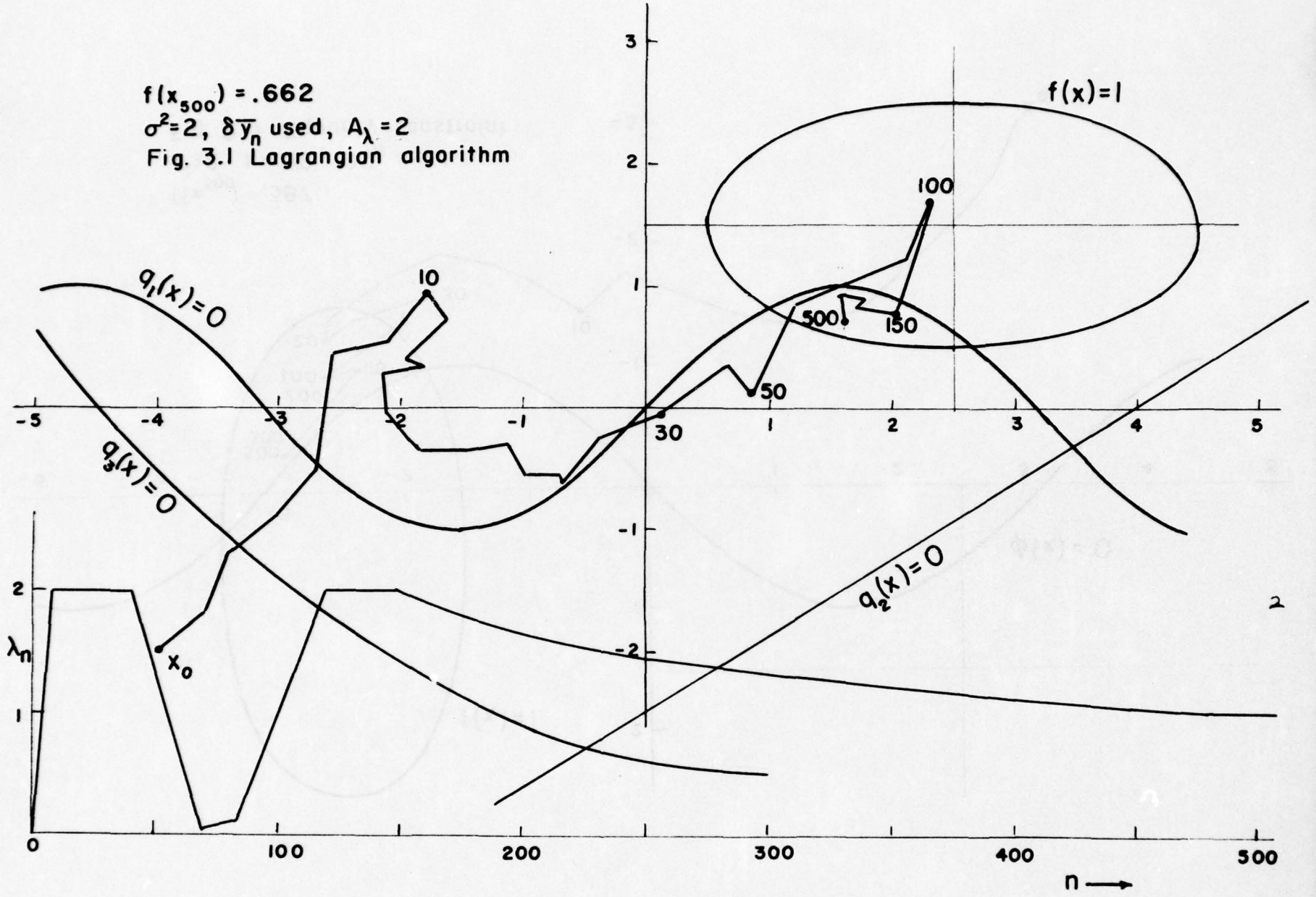


$f(x_{500}) = .295$
$\sigma^2 = 2$, $k = 4$, $\delta \bar{y}_n$ used

Fig. 2.2 Equality constraint



$f(x_{500}) = .387$
$\sigma^2=2$, $k=1/2$, $\delta \bar{y}_n$ used
Fig. 2.3 Equality constraint


$f(x_{500}) = .662$
$\sigma^2=2$, $\delta\overline{y}_n$ used, $A_\lambda = 2$
Fig. 3.1 Lagrangian algorithm
f(x)=1
$g_1(x)=0$
$g_3(x)=0$
$g_2(x)=0$
100
10
500
150
50
30
x_0
λ_n
n
2

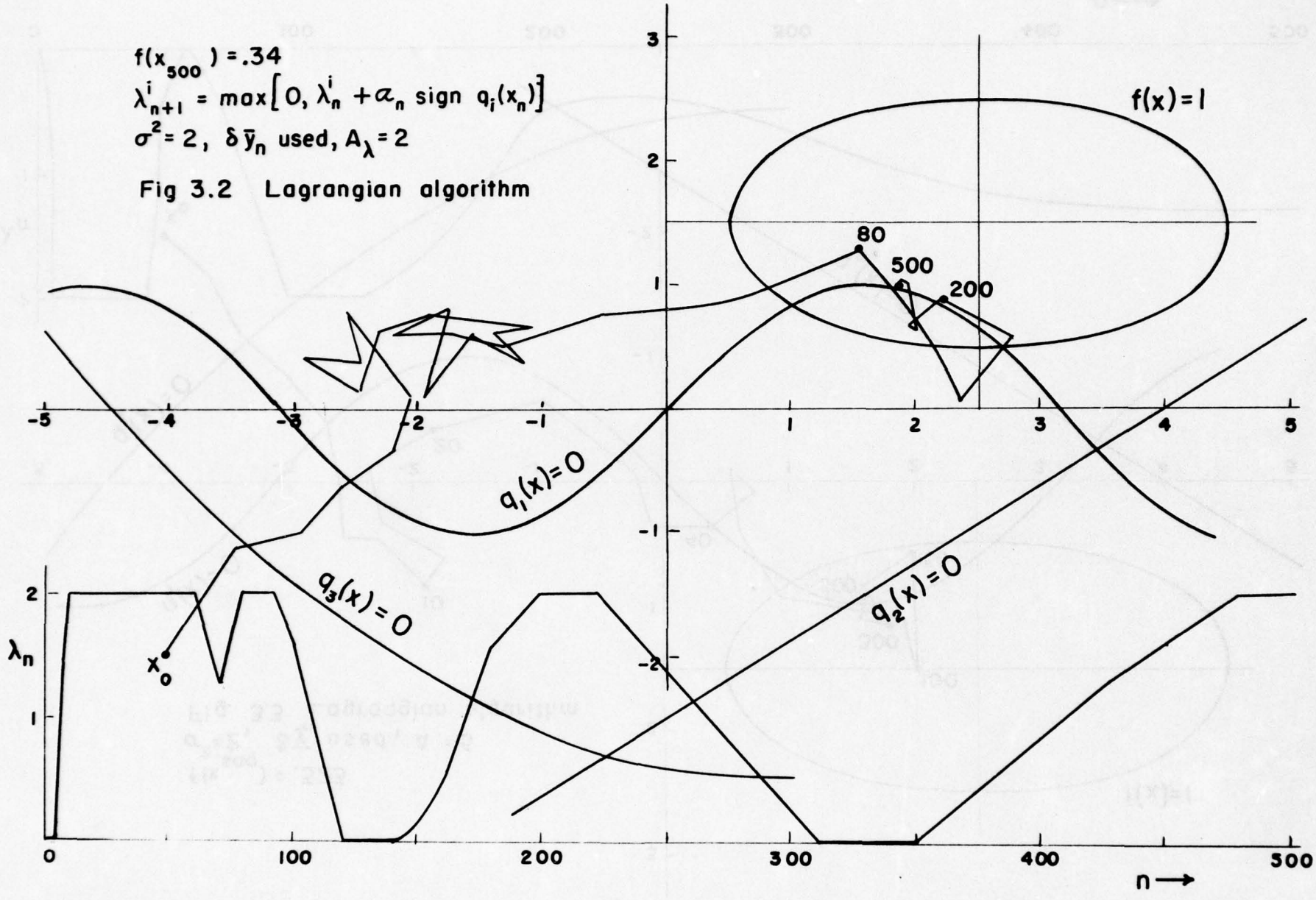
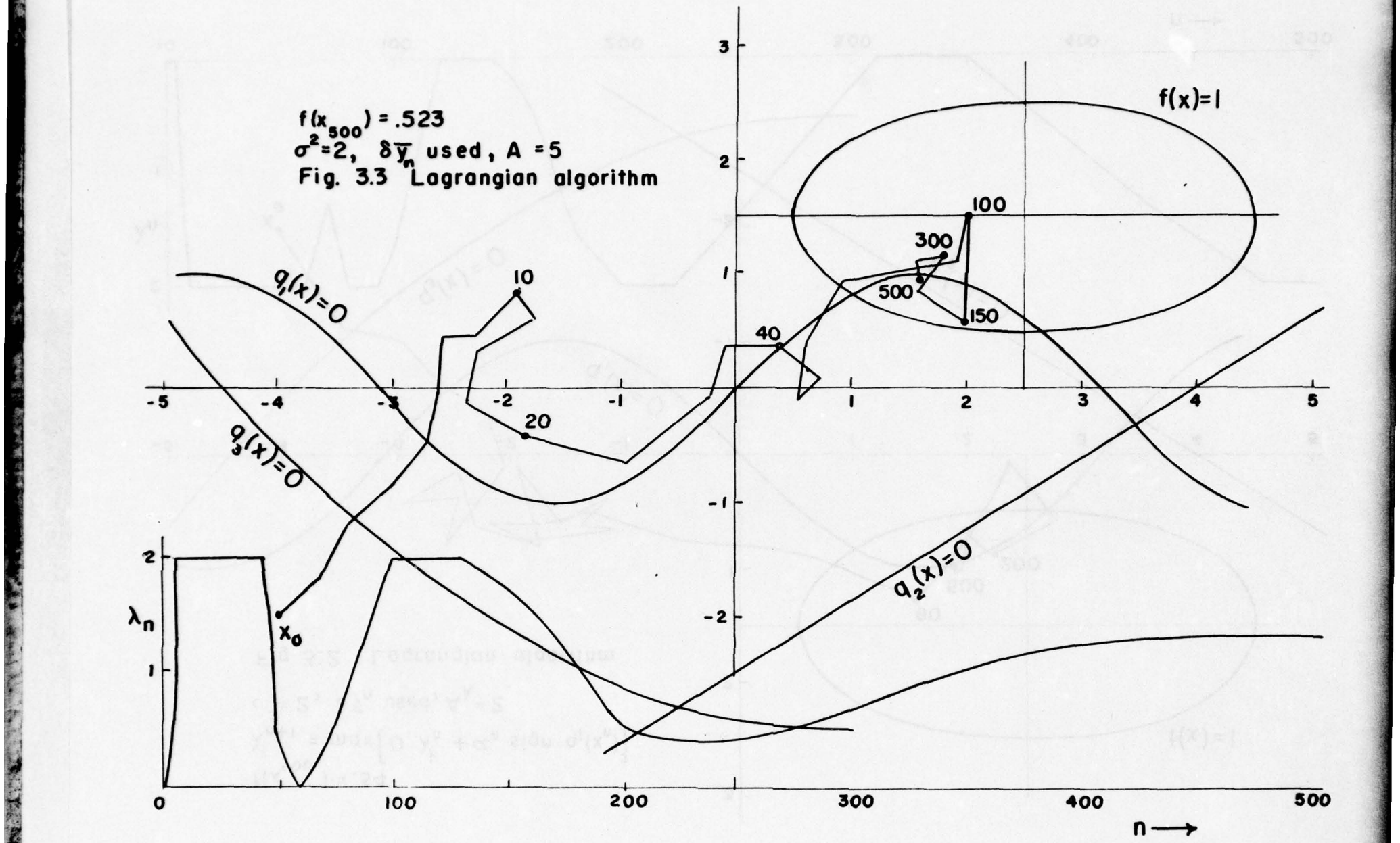
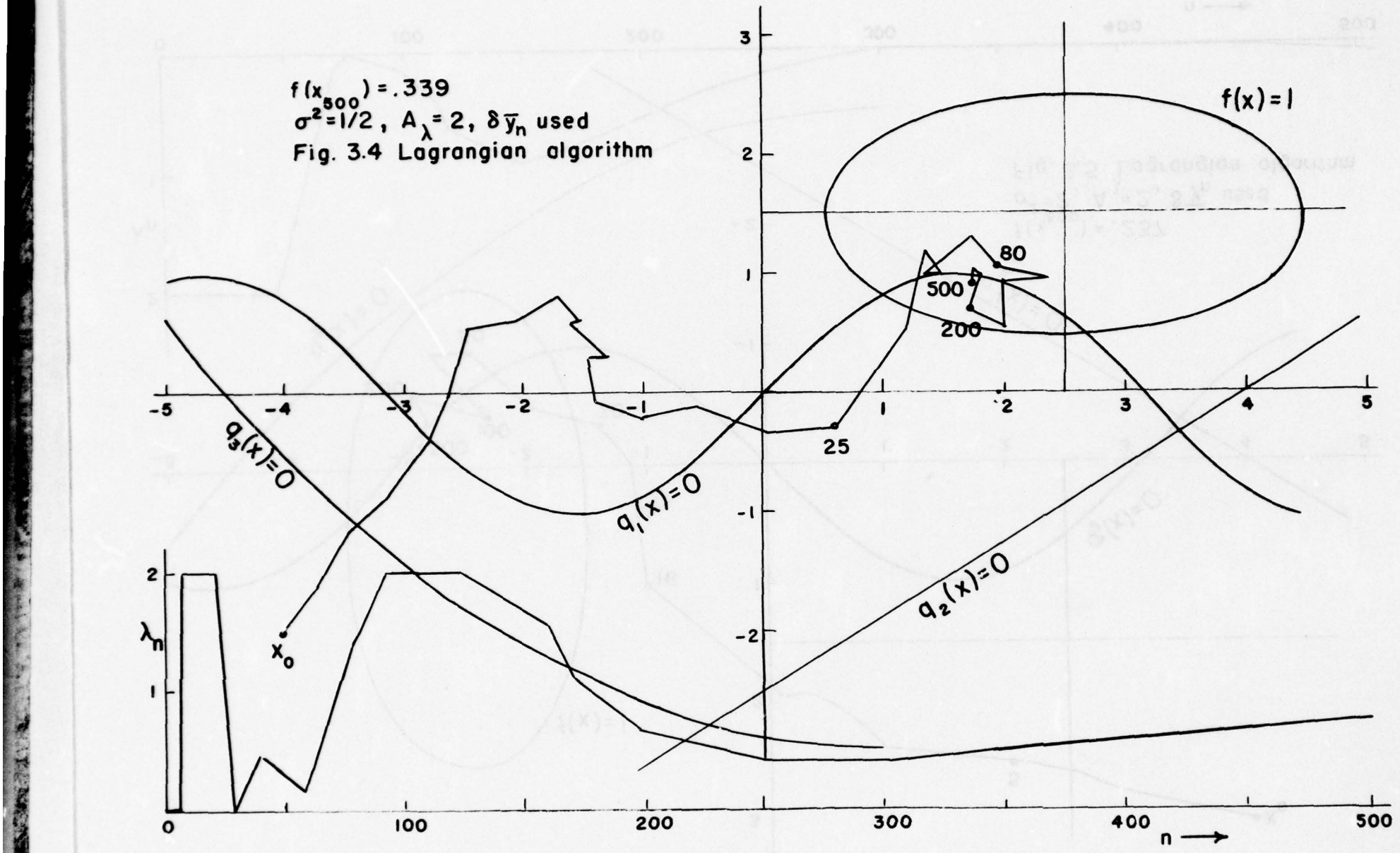


Fig 3.2 Lagrangian algorithm


$f(x_{500}) = .523$
$\sigma^2 = 2$, $\delta \bar{y}_n$ used, A = 5
Fig. 3.3 Lagrangian algorithm
f(x)=1
$q_1(x)=0$
$q_2(x)=0$
$q_3(x)=0$
100
300
500
150
40
10
20
x_0
λ_n
n


$f(x_{500}) = .339$
$\sigma^2 = 1/2$, $A_\lambda = 2$, $\delta \bar{y}_n$ used
Fig. 3.4 Lagrangian algorithm
f(x) = 1
80
500
200
25
x_0
$q_3(x)=0$
$q_1(x)=0$
$q_2(x)=0$
λ_n
n

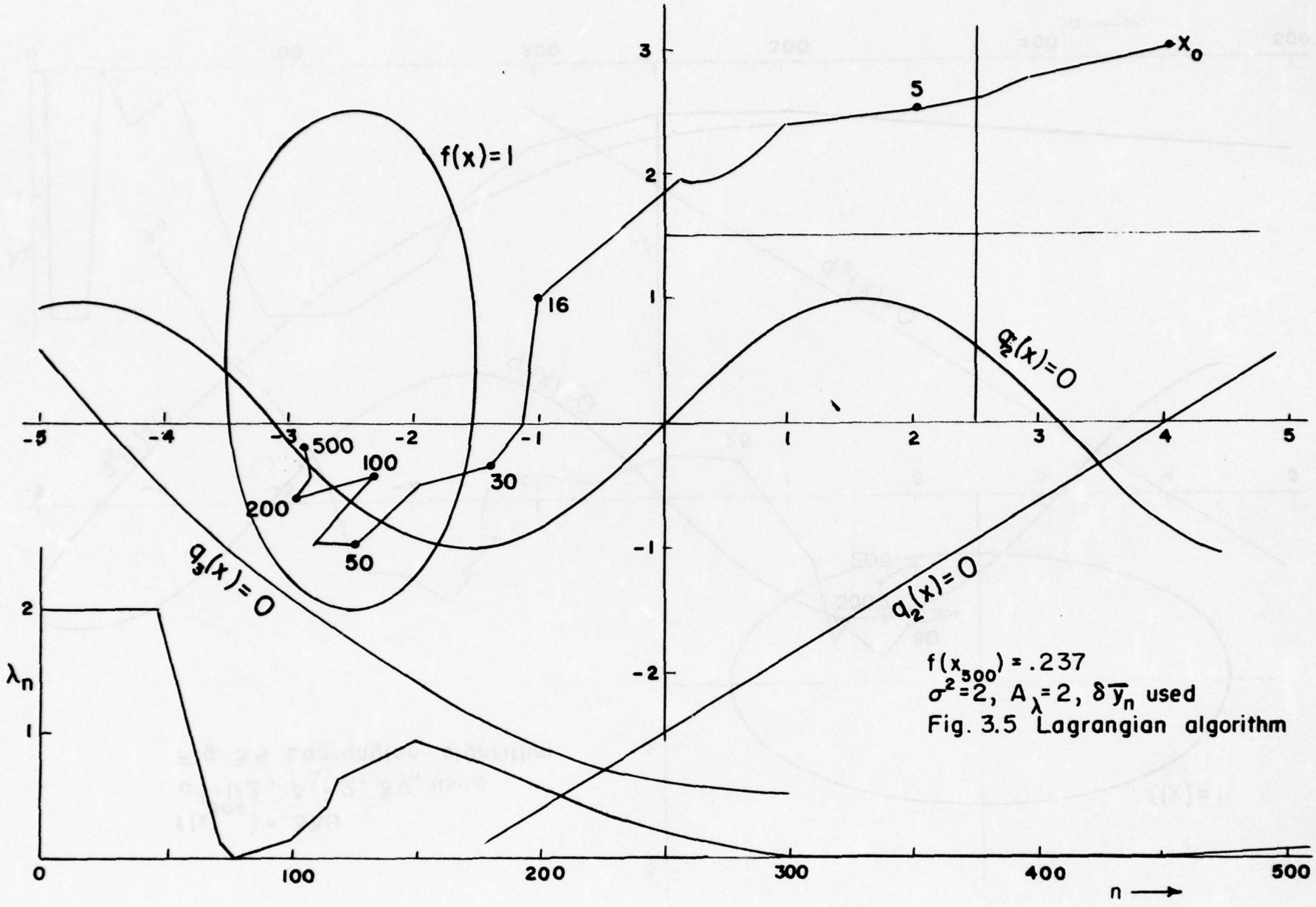
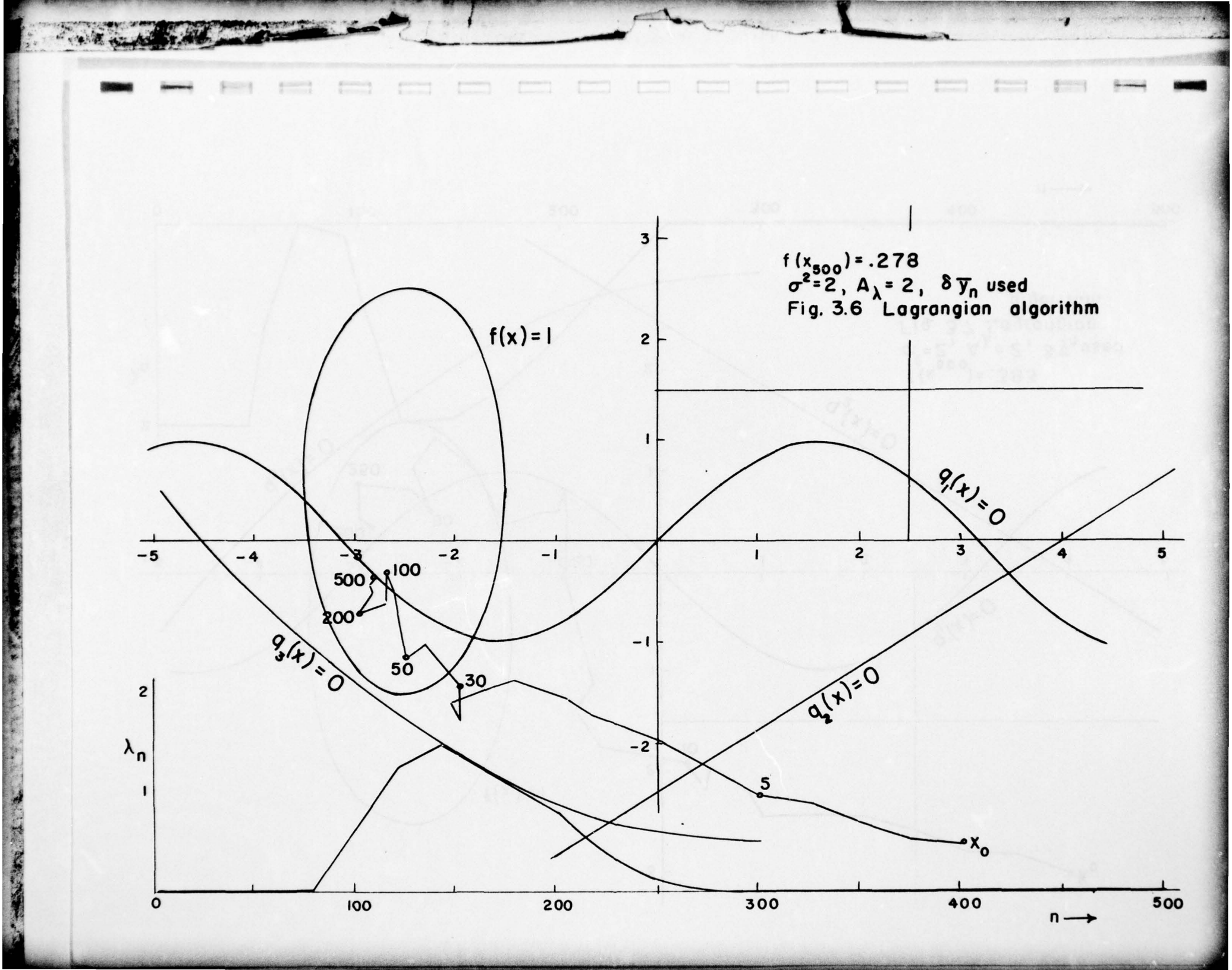
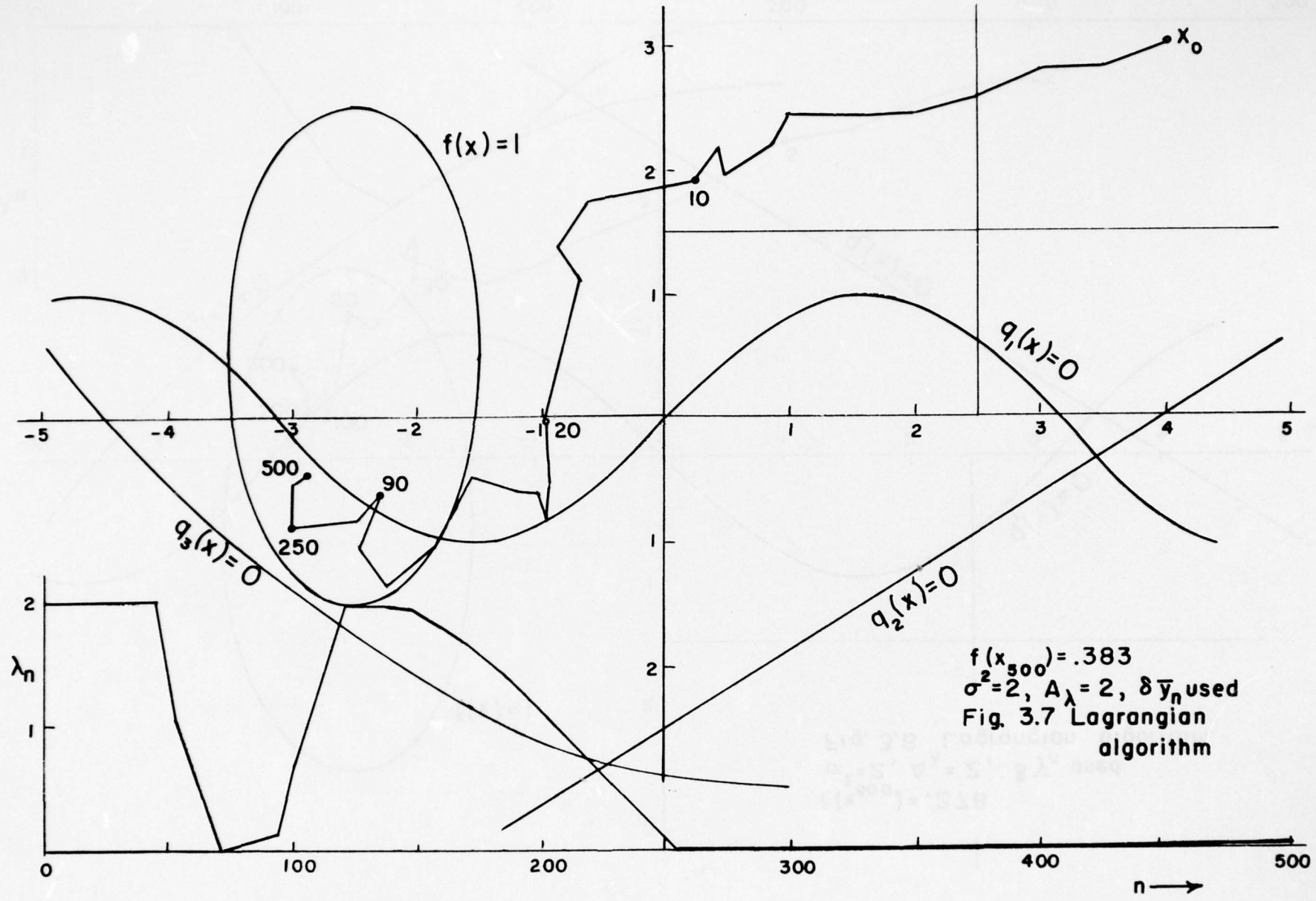
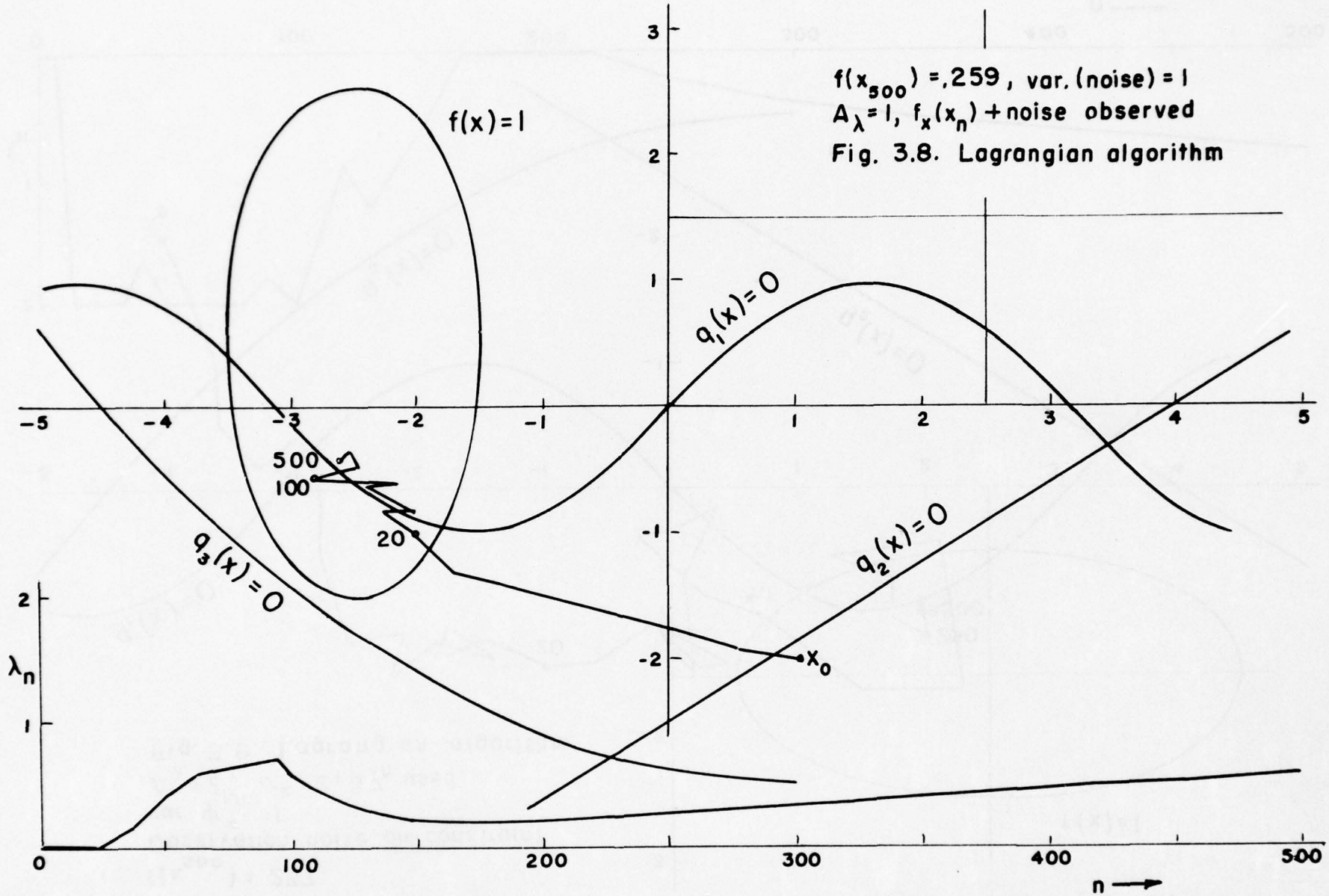


Fig. 3.5 Lagrangian algorithm

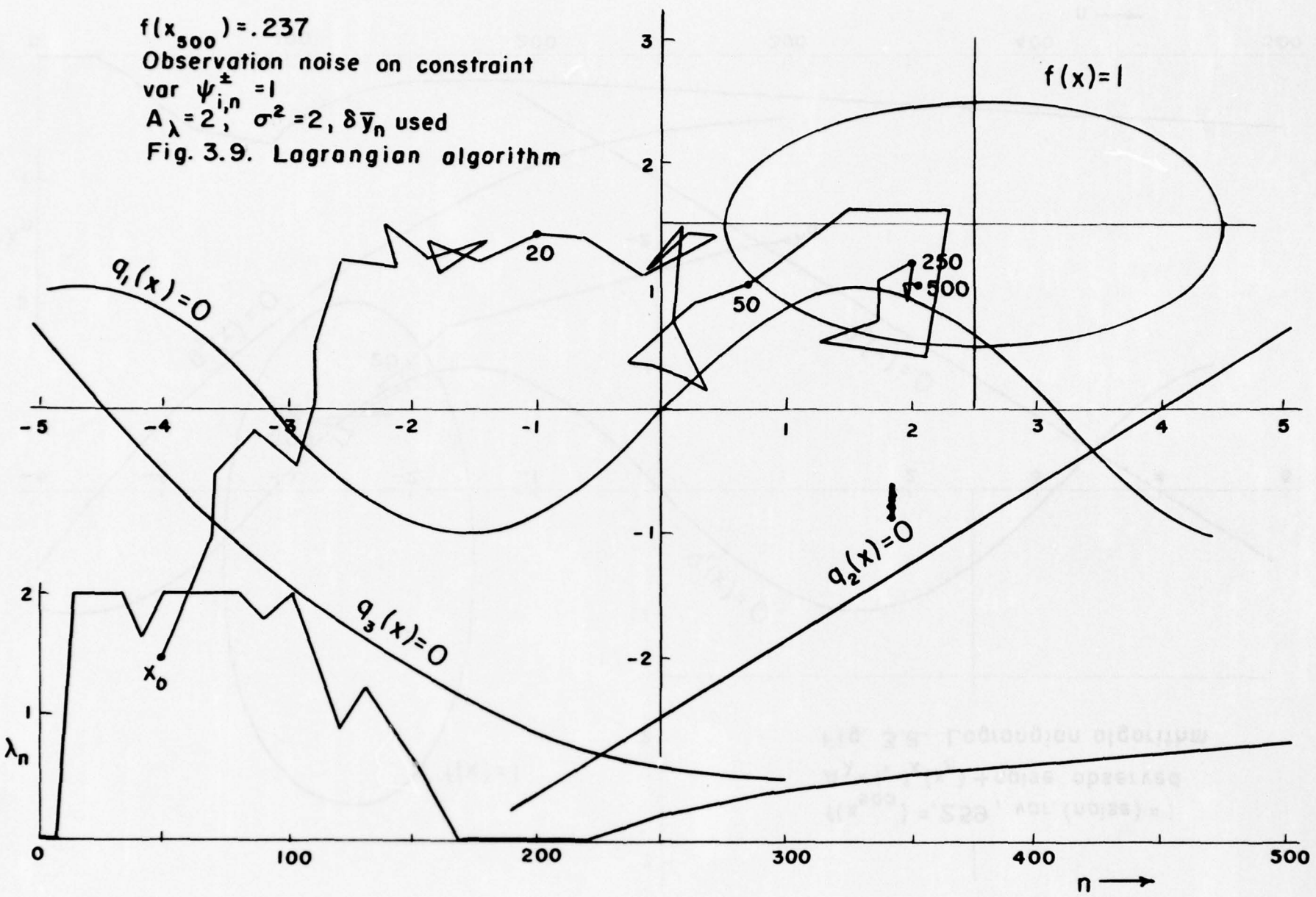

$f(x_{500}) = .278$
$\sigma^2 = 2$, $A_\lambda = 2$, $\delta \bar{y}_n$ used
Fig. 3.6 Lagrangian algorithm
f(x) = 1
$q_1(x) = 0$
$q_2(x) = 0$
$q_3(x) = 0$
500
100
200
50
30
5
x_0
λ_n
n
3
2
1
-1
-2
-3
-4
-5
0
100
200
300
400
500

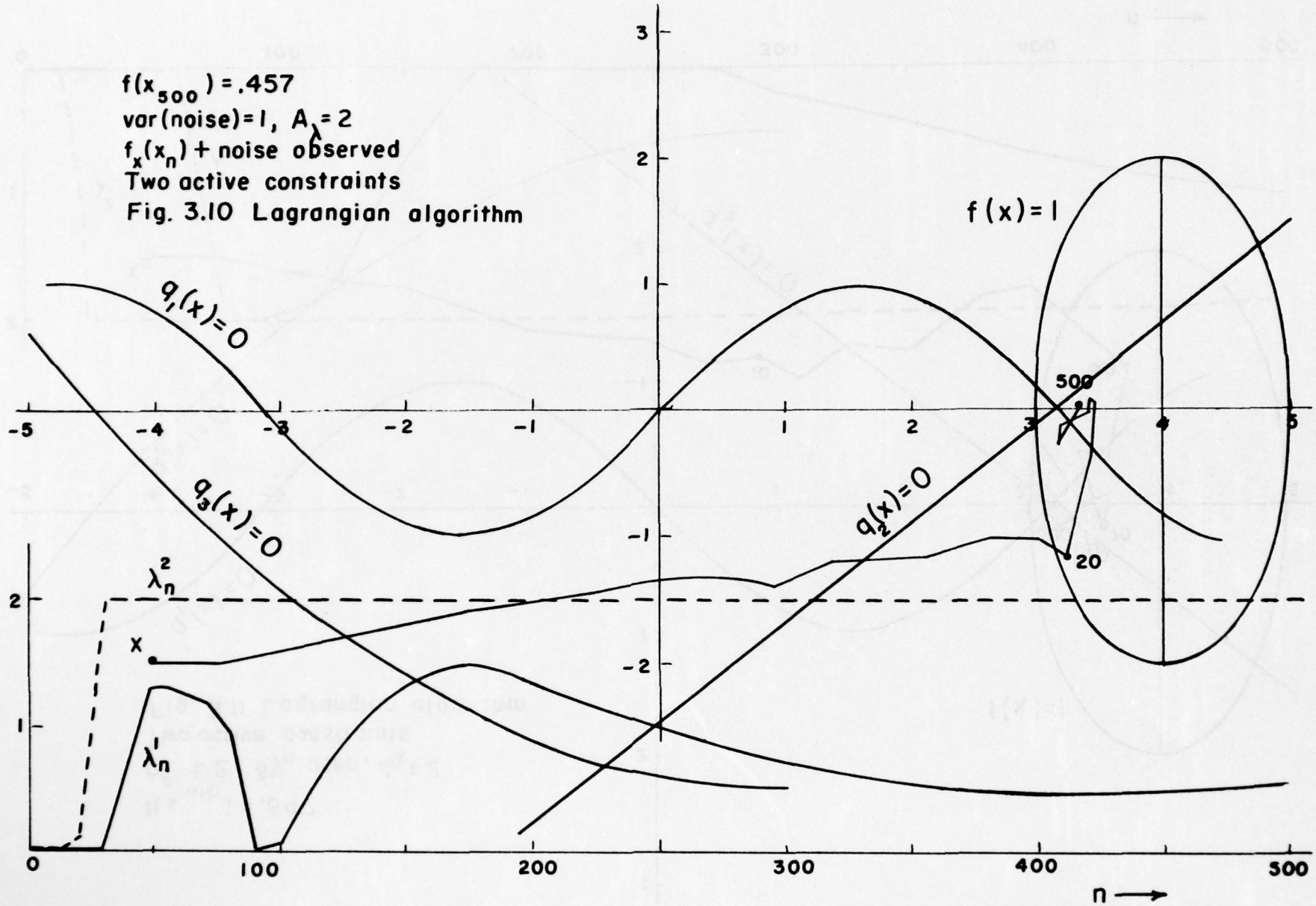

f(x) = 1
x_0
10
$q_1(x)=0$
$q_2(x)=0$
$q_3(x)=0$
500
90
250
$f(x_{500}) = .383$
$\sigma^2=2$, $A_\lambda = 2$, $\delta \bar{y}_n$ used
Fig. 3.7 Lagrangian algorithm
algorithm
λ_n
n

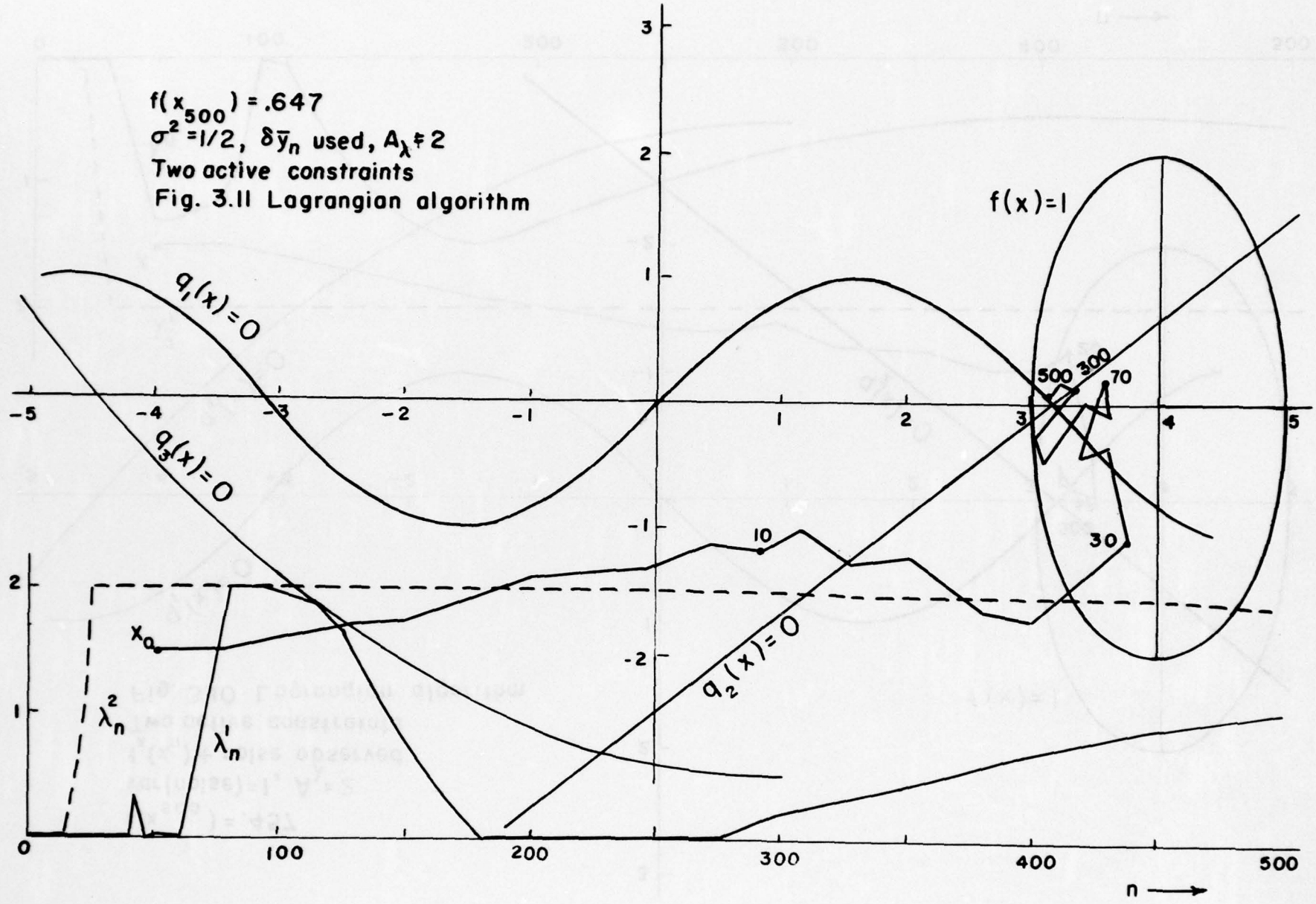


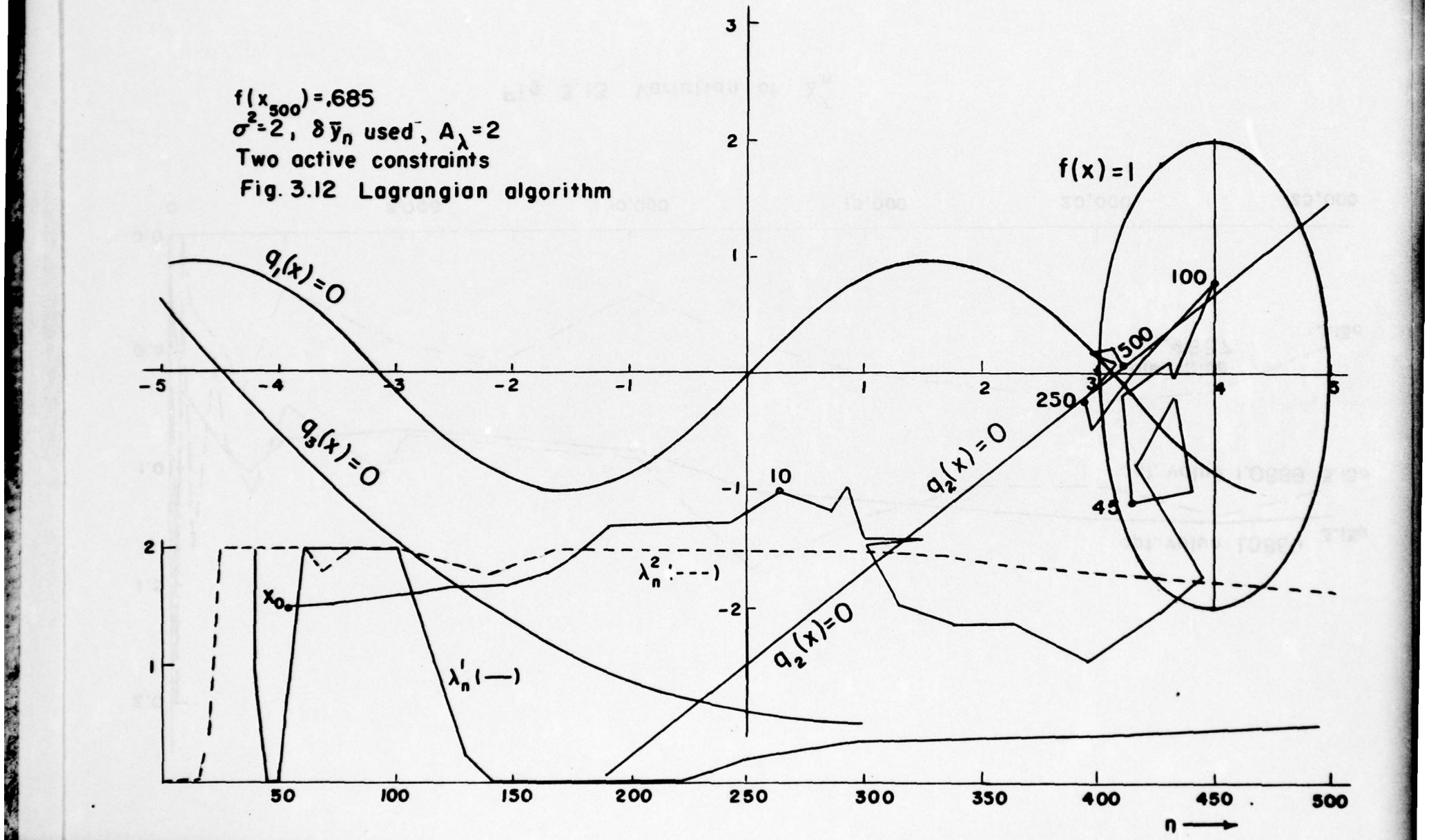
$f(x_{500}) = .259$, var. (noise) = 1
$A_\lambda = 1$, $f_x(x_n)$ + noise observed

Fig. 3.8. Lagrangian algorithm


f(x500) = .237
Observation noise on constraint
var ψ±i,n = 1
Aλ = 2; σ² = 2, δȳn used
Fig. 3.9. Lagrangian algorithm
f(x) = 1
q1(x) = 0
q2(x) = 0
q3(x) = 0
x0
20
50
250
500
λn
n
0
100
200
300
400
500


f(x500) = .457
var(noise) = 1, Aλ = 2
fx(xn) + noise observed
Two active constraints
Fig. 3.10 Lagrangian algorithm
f(x) = 1
q1(x) = 0
q2(x) = 0
q3(x) = 0
500
20
λ²n
λ'n
x
n ⟶


$f(x_{500}) = .647$
$\sigma^2 = 1/2$, $\delta\bar{y}_n$ used, $A_\lambda \neq 2$
Two active constraints
Fig. 3.11 Lagrangian algorithm
f(x)=1
$g_1(x)=0$
$g_3(x)=0$
$g_2(x)=0$
500
300
70
30
10
x_0
λ_n^2
λ_n^1
n


f(x500) = .685
σ² = 2, δȳn used, Aλ = 2
Two active constraints
Fig. 3.12 Lagrangian algorithm
f(x) = 1
q1(x) = 0
q3(x) = 0
q2(x) = 0
q2(x) = 0
100
500
250
10
45
xo
λ²n (---)
λ¹n (—)
η

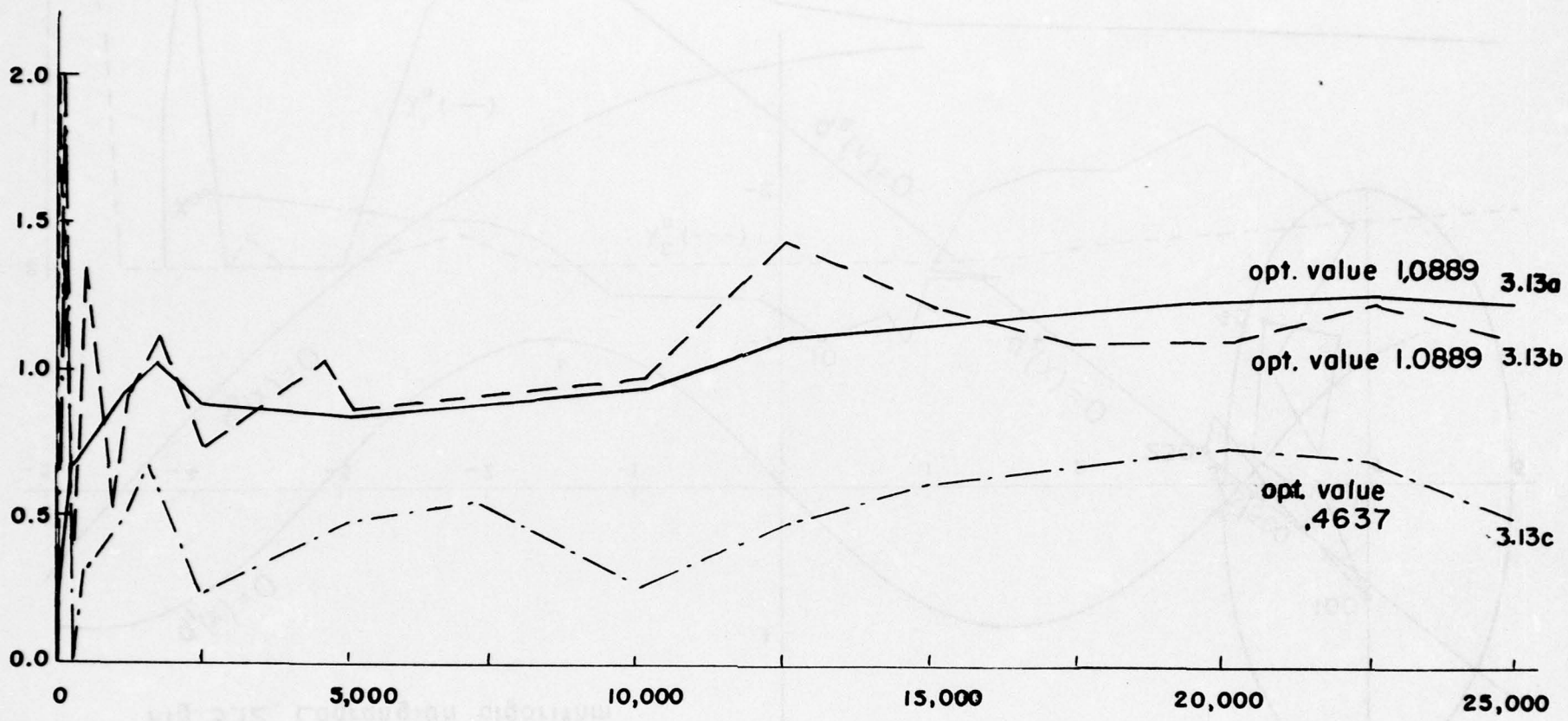
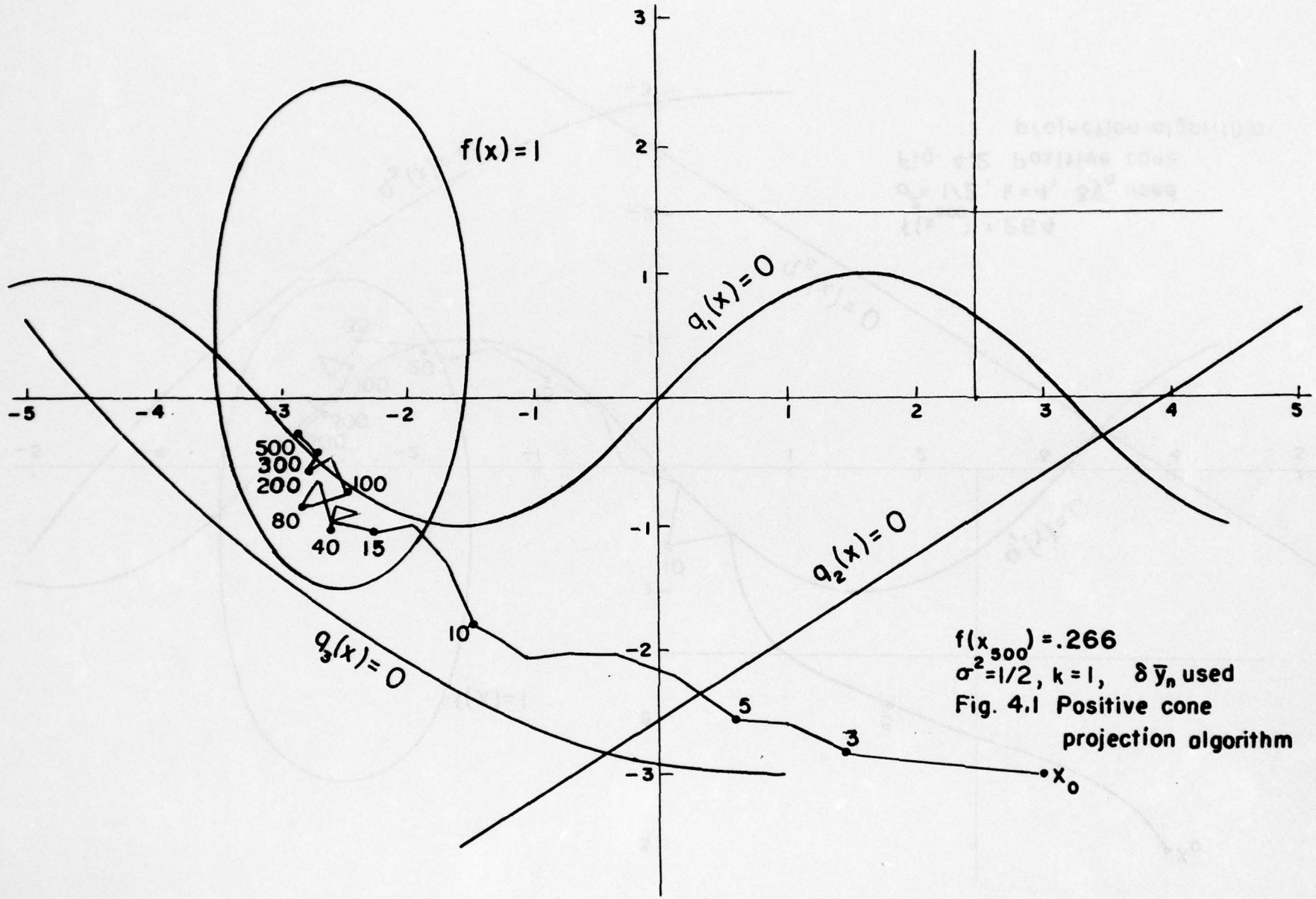
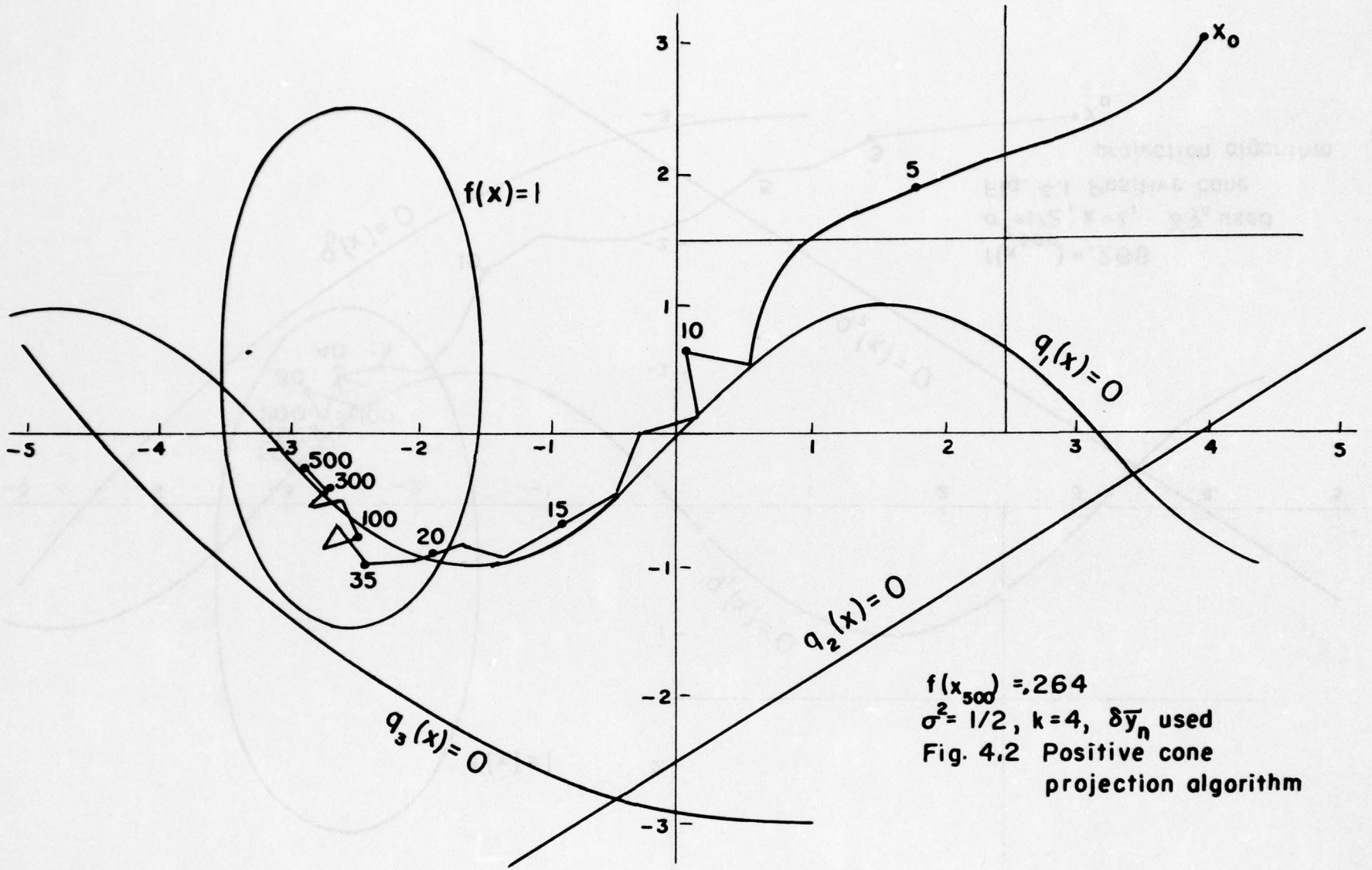


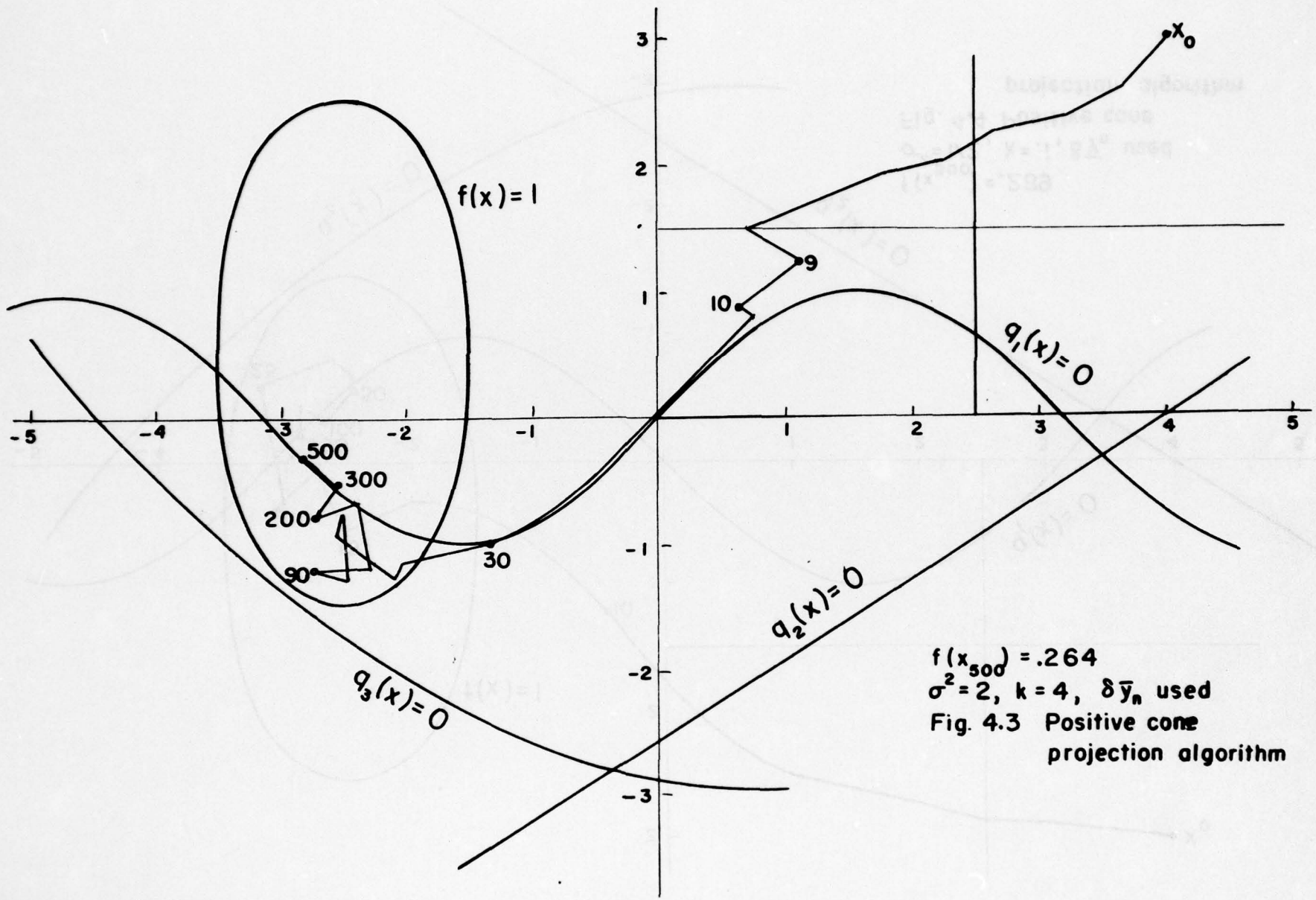
Fig. 3.13 Variation of λ_n'



$f(x_{500}) = .266$
$\sigma^2 = 1/2$, $k = 1$, $\delta \bar{y}_n$ used
Fig. 4.1 Positive cone projection algorithm

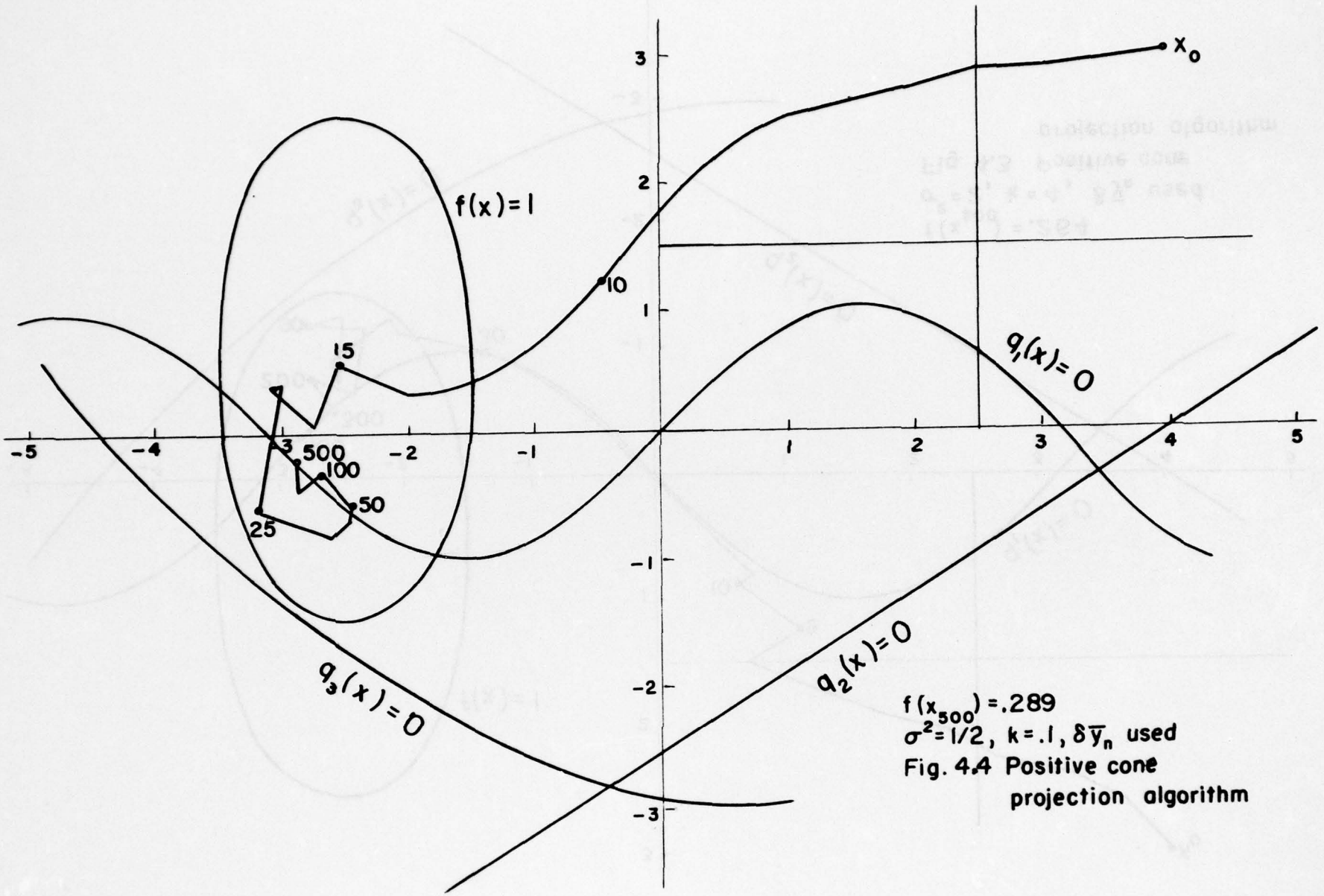


$f(x_{500}) = .264$
$\sigma^2 = 1/2$, $k=4$, $\delta\bar{y}_n$ used
Fig. 4.2 Positive cone projection algorithm



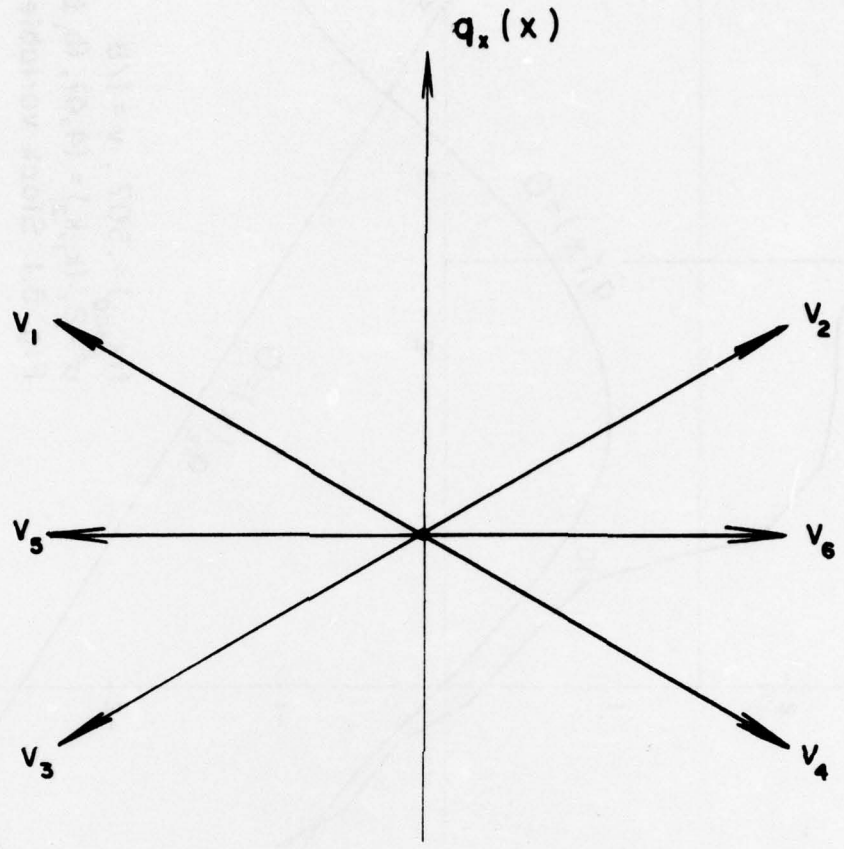
$f(x_{500}) = .264$
$\sigma^2 = 2$, $k = 4$, $\delta \bar{y}_n$ used

Fig. 4.3 Positive cone projection algorithm



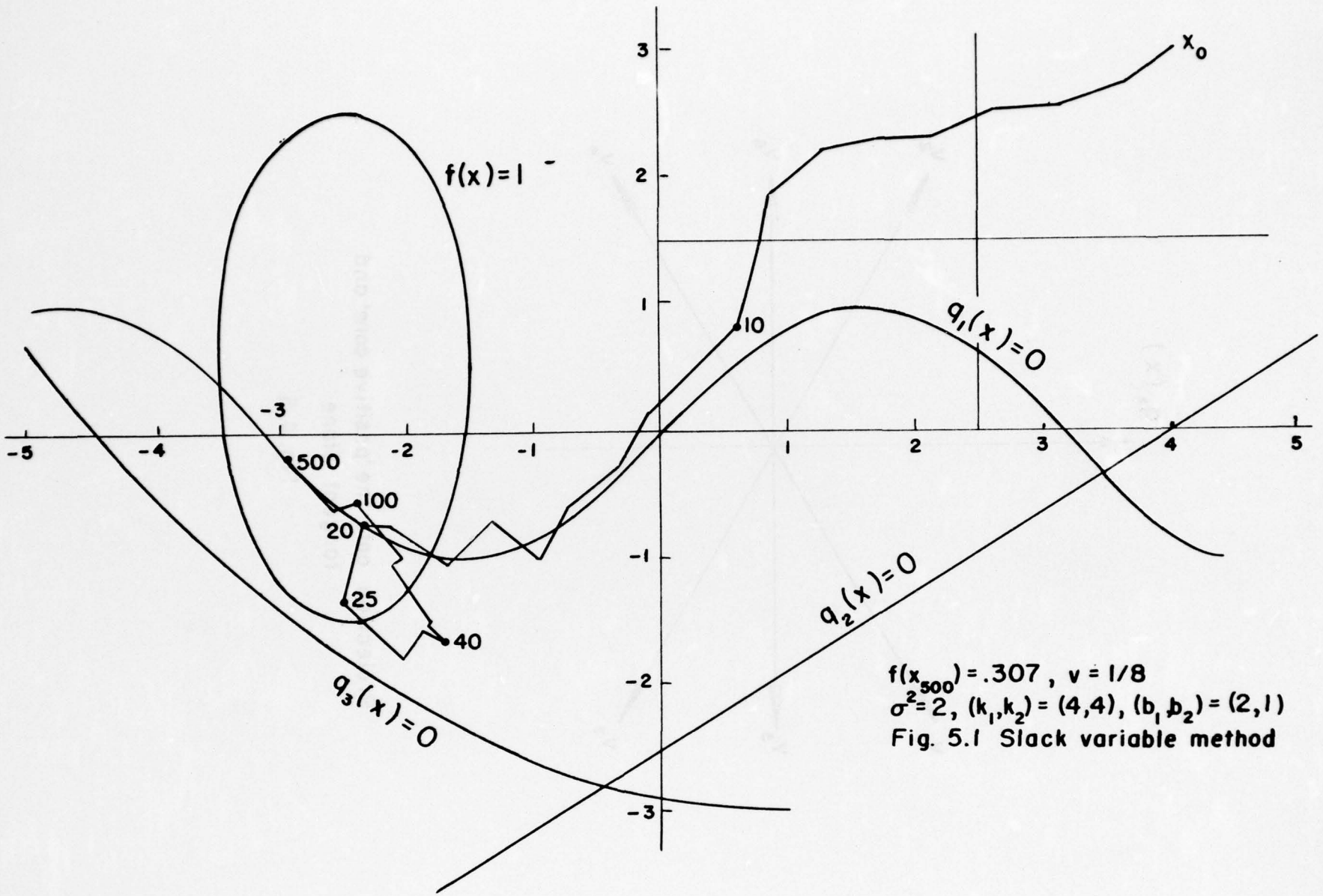
$f(x_{500}) = .289$
$\sigma^2 = 1/2$, $k = .1$, $\delta \overline{y}_n$ used

Fig. 4.4 Positive cone projection algorithm

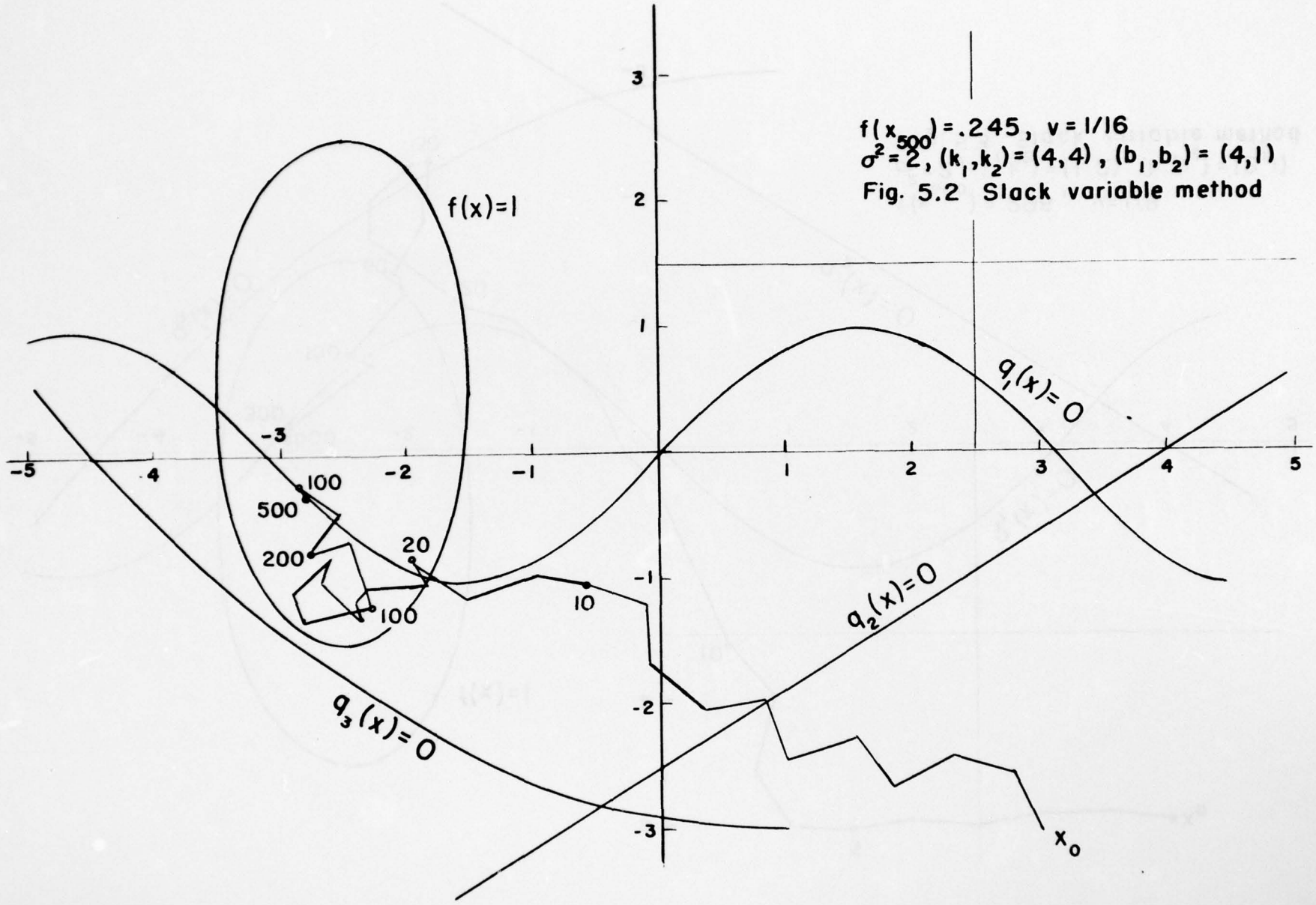


Projection onto the 'positive cone' and tangent plane

Fig. 4.5

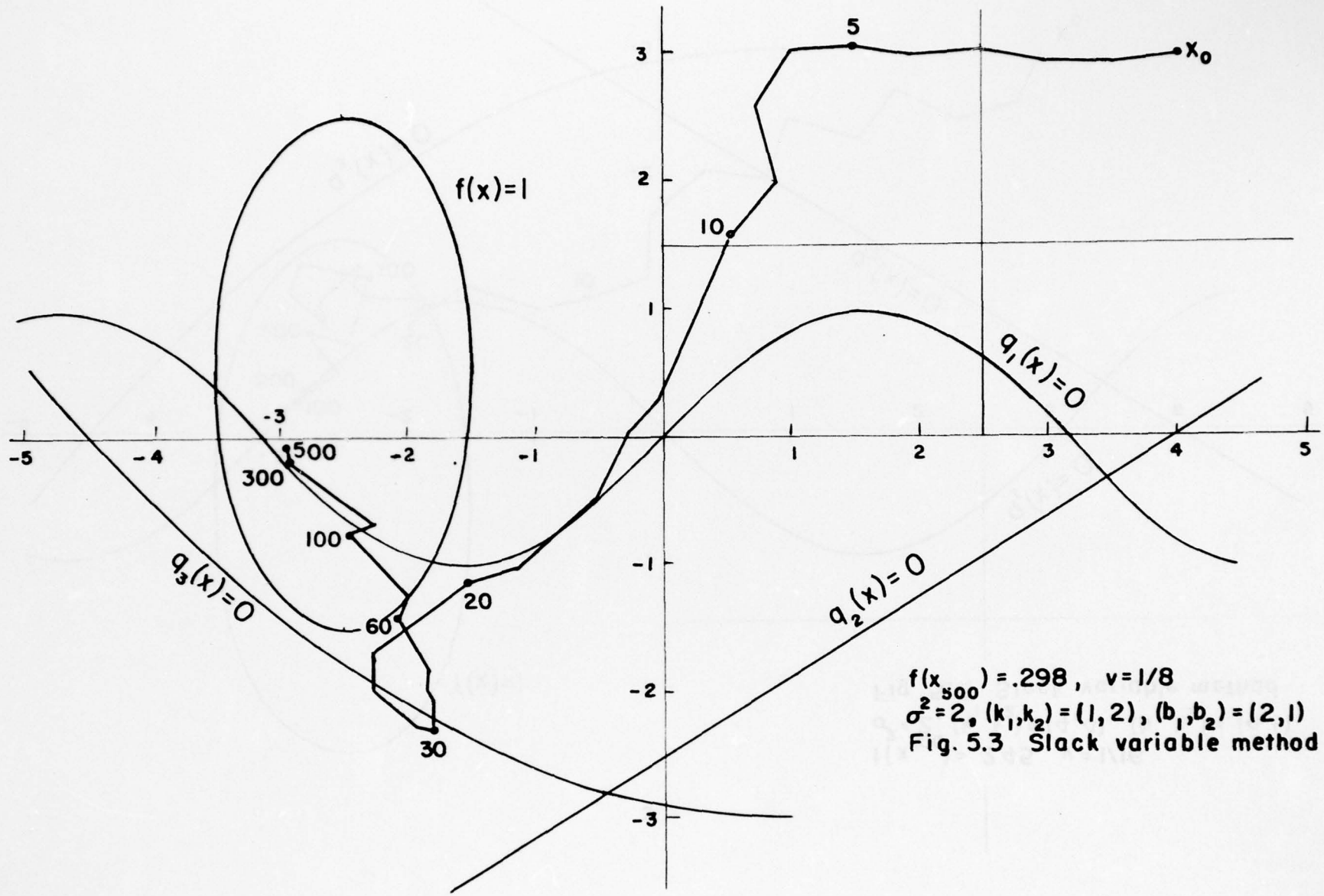


$f(x_{500}) = .307$, $v = 1/8$
$\sigma^2 = 2$, $(k_1, k_2) = (4,4)$, $(b_1, b_2) = (2,1)$
Fig. 5.1 Slack variable method

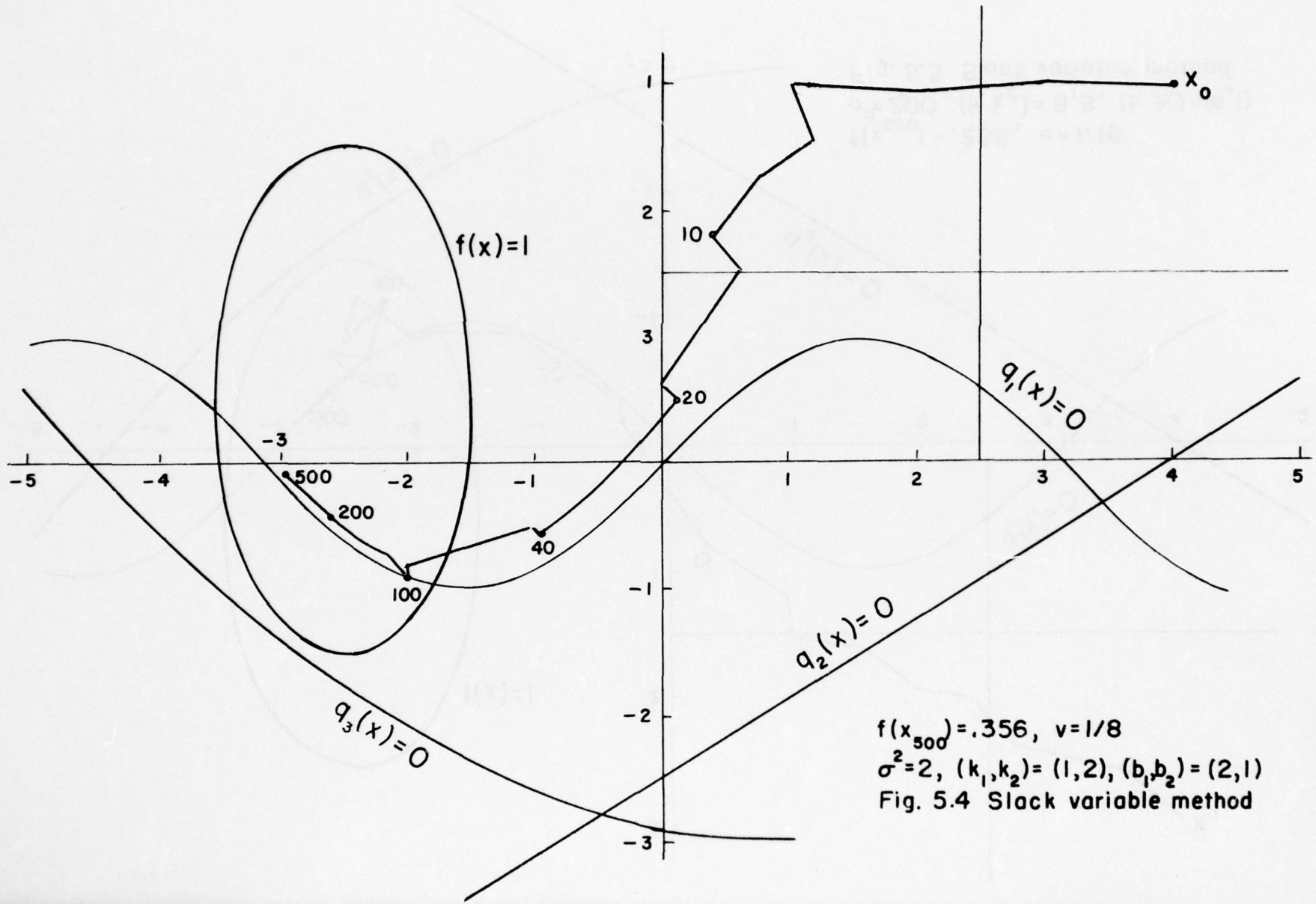


$f(x_{500}) = .245$, $v = 1/16$
$\sigma^2 = 2$, $(k_1, k_2) = (4,4)$, $(b_1, b_2) = (4,1)$

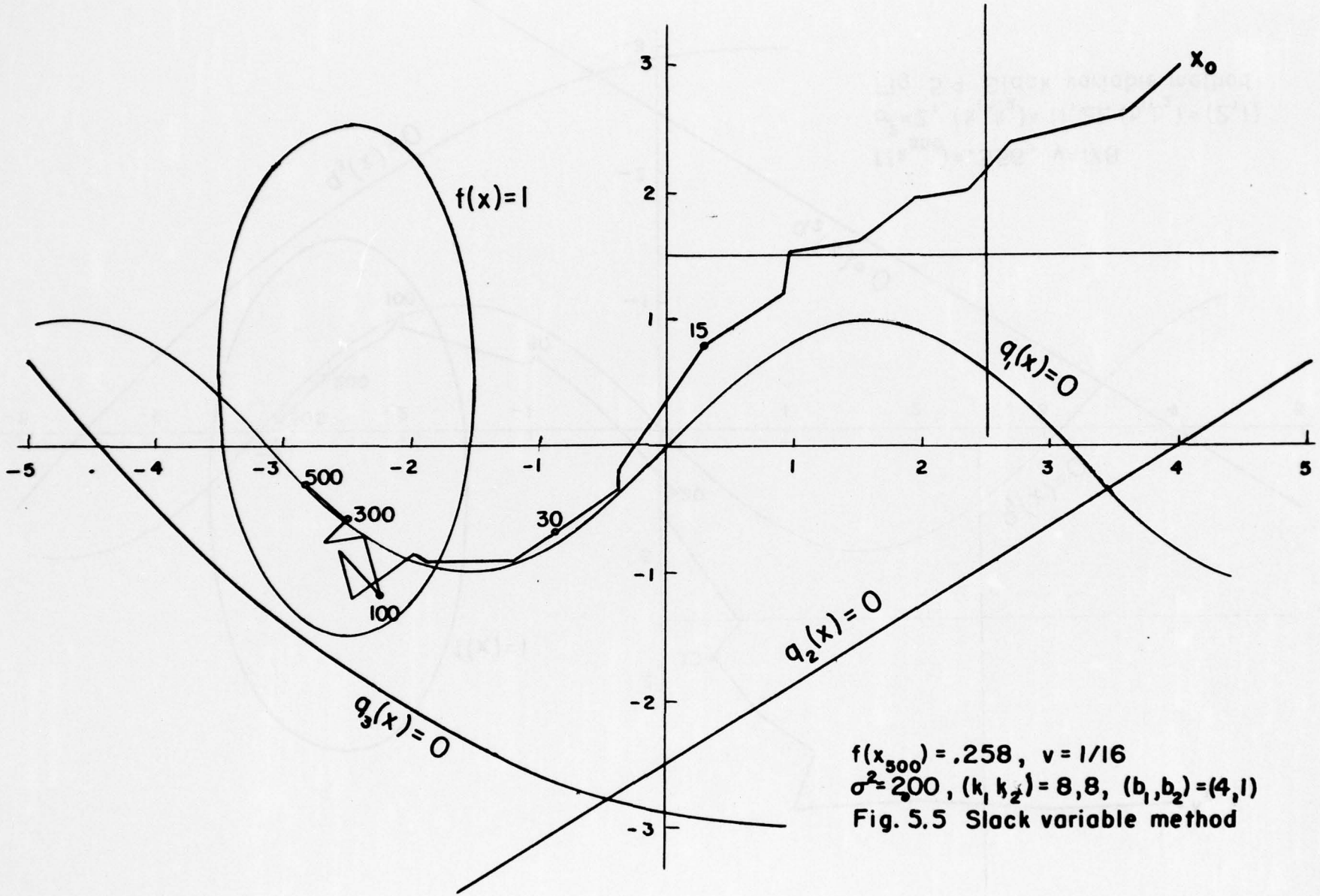
Fig. 5.2 Slack variable method



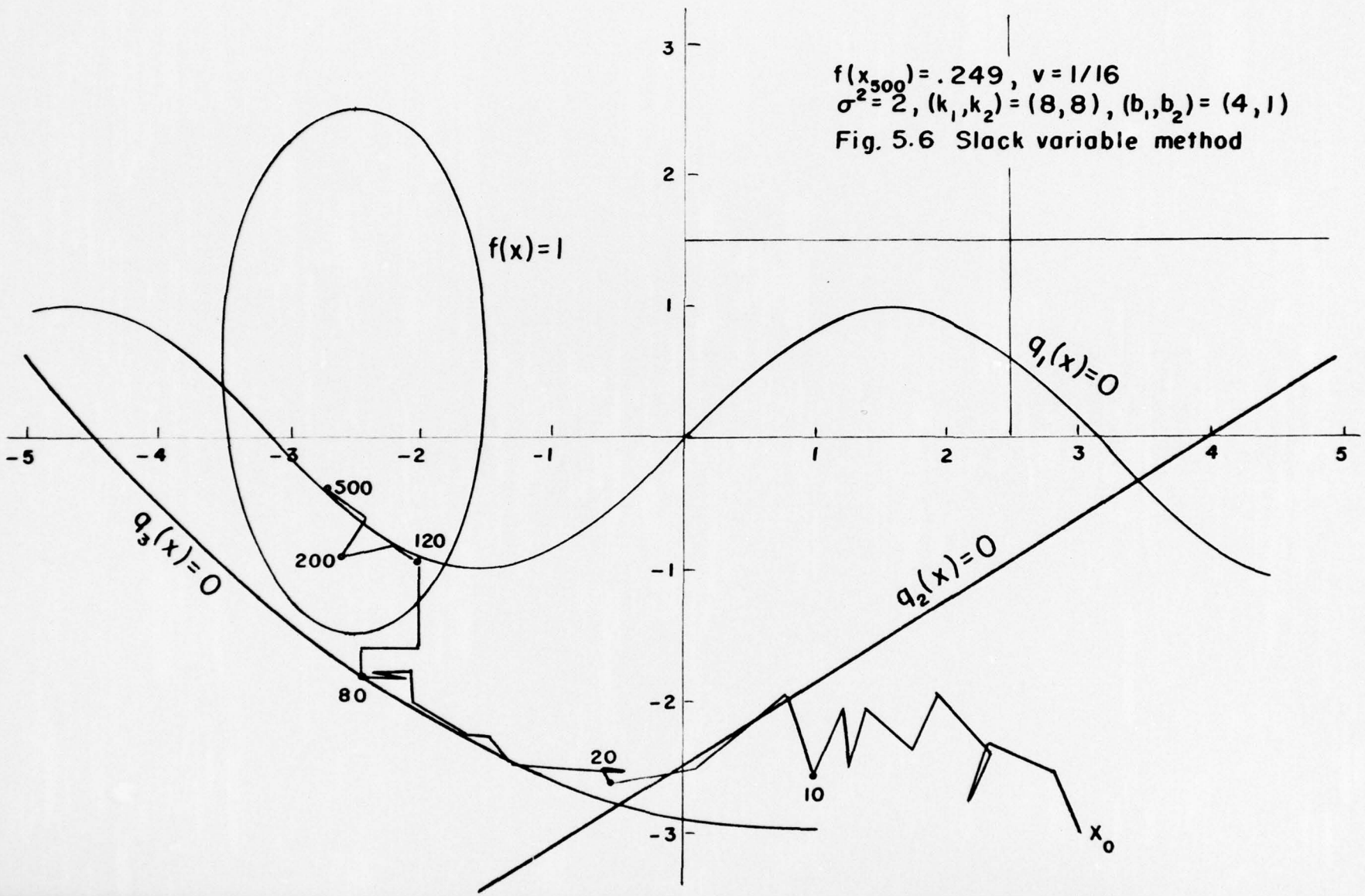
$f(x_{500}) = .298$, $v = 1/8$
$\sigma^2 = 2$, $(k_1, k_2) = (1, 2)$, $(b_1, b_2) = (2, 1)$
Fig. 5.3 Slack variable method



$f(x_{500})=.356$, $v=1/8$
$\sigma^2=2$, $(k_1,k_2)=(1,2)$, $(b_1,b_2)=(2,1)$
Fig. 5.4 Slack variable method



$f(x_{500}) = .258$, $v = 1/16$
$\sigma^2 = 200$, $(k_1, k_2) = 8, 8$, $(b_1, b_2) = (4, 1)$
Fig. 5.5 Slack variable method


$f(x_{500}) = .249$, $v = 1/16$
$\sigma^2 = 2$, $(k_1, k_2) = (8, 8)$, $(b_1, b_2) = (4, 1)$
Fig. 5.6 Slack variable method
f(x)=1
$q_1(x)=0$
$q_2(x)=0$
$q_3(x)=0$
500
200
120
80
20
10
x_0
3
2
1
-1
-2
-3
-5
-4
-3
-2
-1
1
2
3
4
5